Crystal Structure Investigations of Small Heat Shock Proteins and Serine Palmitoyltransferase

WILLIAMSON, IAN,ROBERT

Abstract

Crystal Structure Investigations of Small Heat Shock Proteins and Serine Palmitoyltransferase

Ian R. Williamson

The content of this thesis concerns the pursuit of protein structures from cradle to grave. This process is illustrated though the progress made on two separate proteins. The first is the serine palmitoyltransferase from *Toxoplasma gondii* (TgSPT). This protein is of interest as sphingolipid biosynthesis in apicomplexan parasites is currently the subject of study as a possible drug target, most notably in *Plasmodium*. We begin by looking at the techniques of bioinformatics used to analyse the amino acid sequence of this protein, and by calculation and comparison with other known protein sequences we are able to begin to identify the different domains of the protein and also to begin to identify their respective functions. Bioinformatic techniques are then further used in order to design constructs to be moved forward for overproduction *in vivo*. A number of different constructs showing significant soluble expression of various truncations of TgSPT are reported, and various techniques relating to overexpression an purification are considered, including high-throughput screening work undertaken at the Oxford Protein Production Facility. The second protein looked at is the small heat shock protein from *Methanococcus jannaschii* (MjHSP). This protein is used as a more tractable analogue for the human α-B-crystallin. We look here at the wild type protein as well as a mutant mimicking the disease-causing R120G mutation in α-B-crystallin, R107G, and also two different truncation mutants, mimicking the Q151X truncation. The processes relating to crystal growth and optimisation, to data collection, and to data analysis and structure solution are then considered. Finally, a redetermined structure of wild type MjHSP at significantly increased resolution, and the structure of the R107G mutant are reported.

Crystal Structure Investigations of Small Heat Shock Proteins and Serine Palmitoyltransferase

Ian Robert Williamson

Thesis for the Degree of Master of Philosophy

Department of Chemistry &
Department of Biological and Biomedical Sciences
University of Durham

2012

Contents

Abbreviations

AA – Amino Acid

ADA – N-(2-Acetamido)iminodiacetic Acid

AOS – α-Oxoamine Synthase

BLAST – Basic Logical Alignment Search Tool

CAPS – N-Cyclohexyl-3-aminopropanesulfonic Acid

CCD – Charge Coupled Device

CHES – N-Cyclohexyl-2-aminoethanesulfonic Acid

CHO – Chinese Hamster Ovary

DNA – Deoxyribonucleic acid

dNTP – Deoxynucleotide Phosphate

DLS – Diamond Light Source

DTT – Dithiothreitol

EDTA – Ethylenediaminetetraacetic Acid

EM – Electron Microscopy

EPPS – 3-[4-(2-Hydroxyethyl)-1-piperazinyl]propanesulfonic Acid

FWHM – Full Width at Half Maximum

GST – Glutathione S-Transferase

HEPES – 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic Acid

HSP – Heat Shock Protein

IC_{50} – Half Maximal Inhibitory Concentration

IgG – Immunoglobulin G

IPTG – Isopropyl-β-D-thiogalactopyranoside

LB – Lsogeny Broth

LC FTICR MS – Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

MES – 2-(N-Morpholino)ethanesulfonic Acid

MOPS – 3-(N-Morpholino)propanesulfonic Acid

MPD – 2-Methyl-2,4-pentanediol

NCS – Non-crystallographic Symmetry

Ni-NTA – Nickel-nitriloacetic Acid

NMR – Nuclear Magnetic Resonance Spectroscopy

OD – Optical Density

OPPF – Oxford Protein Production Facility

PCR – Polymerase Chain Reaction

PDB – Protein Databank

pI – Isoelectric Point

PIPPS – Piperazine-1,4-bis(propanesulfonic Acid) PLP – Pyridoxal-5'-phosphate

SDS – Sodium Docecylsulphate

SOC – Super Optimal Broth with Catabolite Repression

PAGE – Polyacrylamide Gel Electrophoresis

rpm – Rotations Per Minute

SPT – Serine palmitoyltransferase

SUMO – Small Ubiquitin-like Modifier

TBE – Tris/Borate/EDTA

TEV – Tobacco Etch Virus

TLC – Thin Layer Chromatography

Tris – Tris(hydroxymethyl)aminomethane

X-gal – 5-Bromo-4-chloro-indolyl-β-D-galactopyranoside

Acknowledgements

Throughout my studies I have relied upon the help and support of many people. Particular thanks go to Simon Padbury and Catherine Bruce for the early help they gave me in gaining a good grounding in molecular biology. I would also like to thank Ian Edwards for his general day to day support, and John Coxon and Andrew Gardner for the contributions they made to this thesis during their time working with me as a biology project student and a summer student respectively. Further acknowledgement goes to Jackie Mosely for her help with mass spectrometry, Louise Bird at the Oxford Protein Production Facility for her support of the work performed there, and Arnaud Baslé, Jon Marles-Wright and Susan Firbank from Newcastle University and all the staff at Diamond Light Source for the support they offered during experiments there. Final thanks go to Ehmke Pohl and Paul Denny for their patient support and guidance throughout this process.

1 The Importance of Structures

Perhaps the most ubiquitous functional unit across all the various forms of life known to science is the protein. A protein is a biological macromolecule made up of one or more chains of amino amino acids, connected and folded in particular ways, giving the protein a unique form and function. The form and function are themselves linked very closely, and a small change in form can create a very significant change in function. Indeed it is in pursuit of a better understanding of the function of these proteins that we seek to gain insight into the details of their structure.

The structure can be broken down into a number of different levels of information. The primary structure, the sequence of amino acids which make up the protein, may be readily extracted from the sequence of DNA which codes for the protein. It is at this level that the smallest changes can have the largest effect. The mutation of one residue out of a thousand may render an active site unable to perform its function, or indeed completely disrupt the higher folding of the protein, again disrupting or even removing its function entirely. Many human diseases can be traced back to the substitution or deletion of one or two amino acids; a change in the primary structure. The secondary structure, the patterns of folds associated with particular bonding motifs between acids, such as α-helices and β-strands, may be to some extent inferred by well known rules of hydrogen bonding motifs and the like. To this end there exist many computer programs dedicated solely to the prediction of secondary structure, working only from the primary structure and prior knowledge of the patterns of acids indicative of bonding motifs. The tertiary structure, the relationship between these substructural elements, may be inferred, to a lesser extent, by analogy with proteins of similar sequence for which the structure has already been determined. The quaternary

structure, the relationship between multiple protein units or chains, is more difficult to access, though may be inferred from behaviour in certain size dependent methods.

One technique of great value in determining the secondary, tertiary, and in some cases quartenary structure of proteins, is X-ray crystallography, and it is this technique that is the main focus of the work to be outlined in the coming chapters. A high resolution crystal structure can give deep insight into the details of these levels of structure for the very simple reason that its end goal is a three dimensional structural model of the protein, giving the location of each amino acid, and in some cases each atom, in the protein. This is however a simplified understanding of the technique, and there are many caveats to be applied, which will be discussed in later chapters.

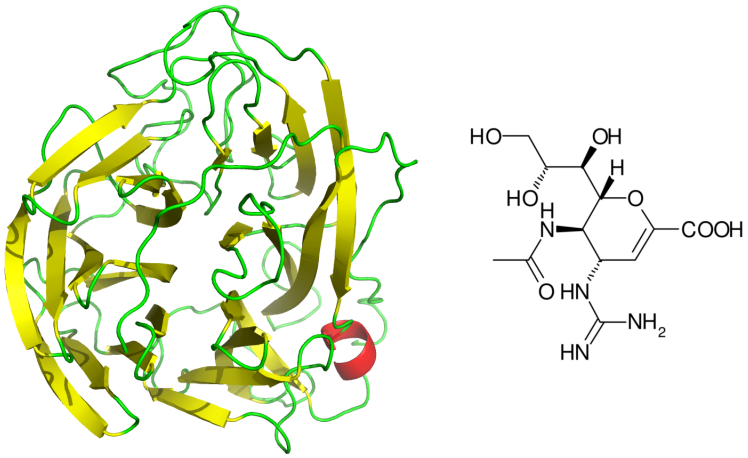


Figure 1: The structure of influenza A N9 neuraminidase (PDB Code: 1NNA) and its inhibitor zanamivir. In the cartoon depiction of the protein α-helix is shown in red; β-sheet in yellow; and random coil in green.

One prominent example where the complete understanding of a protein from primary to quaternary structure played the key role in developing new treatments for human diseases is the N9 neuraminidase from influenza A [1]. Figure 1 shows its X-ray crystal structure, solved to 2.5Å resolution, and the antiviral drug

zanamivir. Neuraminidase is a surface antigen which accounts for 5-10% of the total viral protein and plays a pivotal role in the viral life cycle [2]. Zanamivir was designed as a transition state analogue by rational computer-aided drug design on a basis of the structure of neuraminidase and has proven to be a potent inhibitor, now widely used [3]. Such an example demonstrates quite clearly the use of detailed structural information, and is often the target of a study, but beyond this there are few aspects of a protein's biology which cannot be better understood with full structural information.

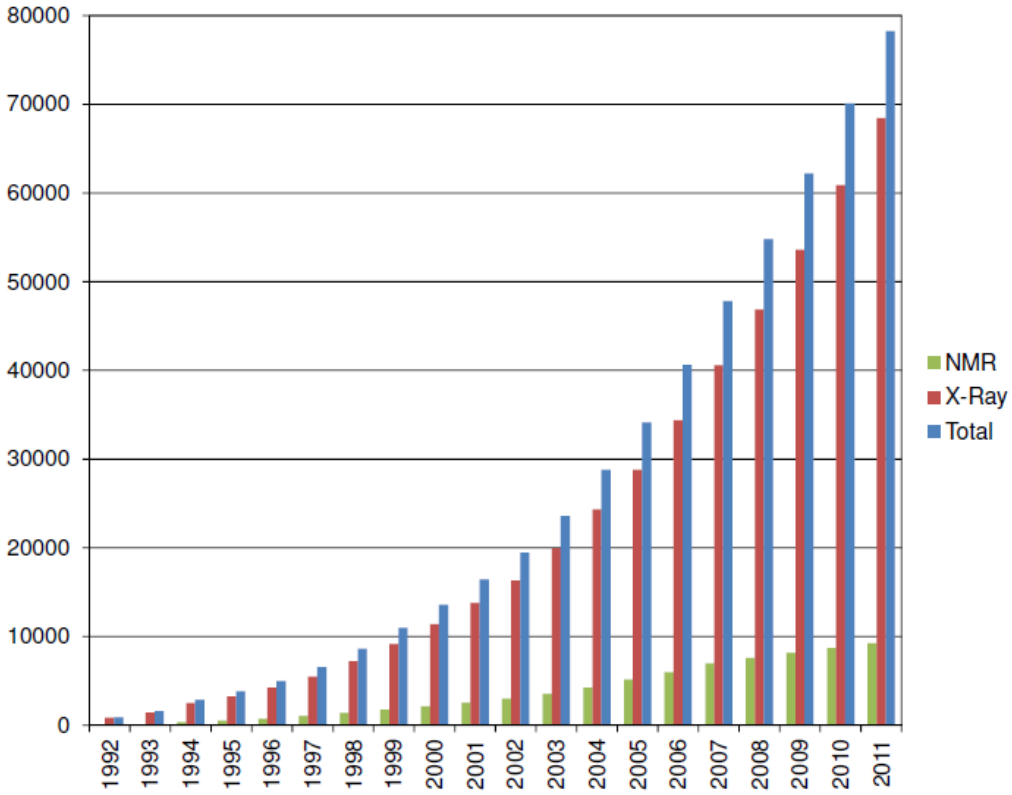


Figure 2: The growth of the total number of structures deposited in the Protein Databank up to the end of 2011, as well as the number which were solved by X-ray crystallography and NMR. EM contributes only 0.5% to the total number of structures. Statistics taken from [4]

1.1 X-Ray Crystallography

X-rays were first observed by Wilhelm C. Röntgen in 1895 [5] who produced the first X-ray image, a picture of his wife's hand. It was Max von Laue who proposed that their wavelength (c.$\sim$1Å)[2] may be of a similar magnitude to the atomic spacing in crystals, and as such first demonstrated their diffraction [6]. Two years later William L. Bragg published the first 3D atomic X-ray crystal structure, that of sodium chloride [7]. Over the following decades, the understanding of the effects involved and the development of modern technologies have driven the science forward. Single crystal X-ray structure determination was first successfully applied to a biological macromolecule by Sir John Kendrew who solved the structure of sperm whale myoglobin in 1958 [8]. There are now nearly 80,000 protein crystal structures recorded in the protein data bank (PDB) (Figure 2), and nearly 600,000 small molecule structures in the Cambridge Structural Database [9]. Indeed protein crystallography has become such a widely used technique that the number of structures in the PDB is now growing exponentially year on year, and with continued investment in large-scale central facilities such as the Diamond Light Source in Oxfordshire, this trend will only continue over the coming years.

1.2 The Serine Palmitoyltransferase from *Toxoplasma gondii*

The work described in this thesis concerns the pursuit of the structures, and the relationship between the structure and function of two separate proteins. The first of these is the Serine Palmitoyltransferase (SPT) from *Toxoplasma gondii*. *T. gondii* is an organism of interest both in itself as a human pathogen of some significance, and more importantly as a more tractable analogue of *Plasmodium*, the causative agent of the disease malaria.

[2] 1 Å $= 1 \times 10^{-10}$ m. By convention, the ångström is the unit of distance used in crystallography due to its convenient order of magnitude relative to the length of a chemical bond.

Our work considers the pathway for sphingolipid biosynthesis as a possible target drugs, and our previous work has shown that inhibition of this pathway causes cell mortality *in vivo* [10]. Within this pathway, this project focuses on the SPT as a key target enzyme in this chain. The main focus of the work on the SPT is in the early stages of structure determination. Various truncates are designed using bioinformatic techniques, and targeted with the aim of isolating significant quantities of pure, monodisperse protein, which if successfully produced may then be taken forward to crystallisation trials. Beyond this, the project aims to produce crystals, from which the structure may be determined, and eventually used to develop novel inhibitors.

1.3 Heat Shock Protein 16.5 from *Methanococcus jannaschii*

The second of the proteins considered is the Heat Shock Protein 16.5 from *Methanococcus jannaschii*. This is again used as an analogue more suitable for laboratory studies, in this case for human αB-crystallin. Human αB-crystallin is itself of interest due to its role in human disease, most particularly myopathy and cataracts. A number of mutations have been identified in human αB-crystallin which have been shown to have a role in causing these conditions, and it is particularly the analogues of these in Heat Shock Protein 16.5 that we will look into.

This part of the work begins with purified protein, which has been isolated by colleagues in Durham University's Department of Biological and Biomedical Sciences, and therefore focuses on the work of crystallisation, diffraction and structure solution. Both the native protein and three mutants, two truncation mutants and one point mutation, are studied. The end goal is to identify what effect the mutations have on the structure, to identify how this shapes their role in human disease, and whether potentially any drug might be developed to mitigate their effect.

2 Sphingolipid Biosynthesis in *Toxoplasma gondii* as a Drug Target

2.1 *Toxoplasma gondii* and Toxoplasmosis

Toxoplasma gondii is an apicomplexan parasite closely related to species such as *Cryptosporidium* and the causative agent of malaria, *Plasmodium* [11]. Due to its tractability in the laboratory *T. gondii* is widely regarded as the model system for the study of such parasites and is therefore well understood. *T. gondii* infection affects a significant proportion of the world's population; it is estimated that 16–40% of people in the UK and USA and 50–80% of people in continental Europe are infected [12]. In general no significant problem arises due to infection, however, if a mother becomes infected for the first time during pregnancy the fetus is at significant risk of retinochoroiditis, hydrocephalus, convulsions and cerebral calcification [13]. In the least severe cases eyesight is impaired, the most severe cases lead to death. The socioeconomic impact of toxoplasmosis is most apparent in terms of the significant costs incurred in caring for children suffering from mental retardation and blindness [14]. Further risks exist to transplant patients treated with immunosuppressive agents and toxoplasmosis is one of the highest causes of death in patients with AIDS with up to 30% of patients in Europe dying from toxoplasmosis [15].

T. gondii has a complicated life cycle, with many stages. Indeed it may only reproduce sexually in cats or other Felidæ. Cat fæces infected with oocysts may then contaminate food and drink passing the infection on to humans directly or via undercooked meat from an animal infected with *T. gondii*. The largest outbreak in humans recorded was linked to a contaminated municipal water reservoir in British Columbia, Canada [16]. Infection via food may be effectively prevented

by a combination of thorough food preparation hygiene and either heating meat to 67℃ [17], cooling it to -13℃ [18] or by gamma irradiation [19]. While current therapies exist, they are generally based on a cocktail of drugs and evidence of limited tolerance in individual cases and rising antibiotic resistance call for renewed efforts in drug development.

2.2 Sphingolipids

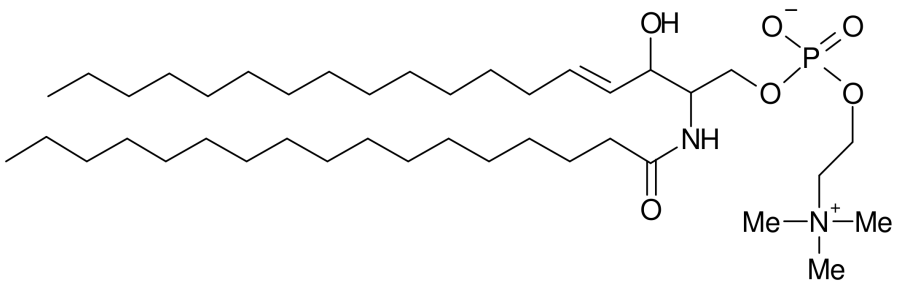


Figure 3: The sphingolipid sphingomyelin. Three main parts may be identified: sphingosine (the bottom long chain), phosphocholine (the right hand group) and a fatty acid (the top long chain).

Sphingolipids are key components of all membranes. They are a class of lipids made up of the aliphatic amino alcohol sphingosine and a fatty acid, for instance sphingosine (figure 3). They were first noted as part of the array of molecules present in cells by J. L. W. Thudicum who noted their presence in the brain in 1884 [20] and were first successfully characterised, gaining their current name, in the middle of the last century [21]. It is now recognised that this class of amino alcohols numbers in the hundreds, with significant variety in all aspects of their chemistry [22]. Their roles in the cell are widespread, including the formation of membrane lipid rafts particularly for cell signalling [23, 24] and recognition [25]. Significant study has been undertaken on the various enzymes involved in this process [26], and recent work has identified sphingolipid biosynthesis as a possible

drug target in apicomplexan parasite species due to the significant differences between the synthetic pathways of animals and other species [10].

2.3 Serine Palmitoyltransferase

Serine palmitoyltransferase (SPT) is an enzyme involved in the early stages of sphingolipid biosynthesis [27, 28]. It catalyses the Claisen condensation of L-serine with palmitoyl CoA, both basic cellular metabolites, to form ketodihydrosphingosine (KDS), the 'long chain base' of sphingolipids [29]. It belongs to the α-oxoamine synthase (AOS) family of enzymes which catalyse condensation reactions between a specific amino acid and an acyl CoA thioester substrate. These enzymes fall into the larger category of fold type I pyridoxal-5'-phosphate (PLP) dependent enzymes [30]. The structures for a number of enzymes from the AOS family have been solved including 8-amino-7-oxononanoate synthase from *E. coli*, which is involved in biotin synthesis [31, 32], 5-aminolevulinate synthase from *Rhodobacter capsulatus* involved in tetrapyrrole metabolism [33, 34], 2-amino-3-ketobutyrate CoA ligase from *E. coli* which is involved in threonine biosynthesis [35] as well as the structure of the SPT from *Sphingomonas paucimobilis* [36].

After SPT, the KDS is reduced to sphinganine by 3-ketosphinganine reductase, which is in turn acylated by dihydroceramide synthase to produce dihydroceramide. This is then again reduced to ceramide by dihydroceramide reductase. This is then transported to the Golgi apparatus where it is further metabolised by sphingolipid synthase, producing phosphosingolipid. In this process, SPT has been identified as the rate determining enzyme [27, 28], and it is this, combined with its early place in the chain of sphingolipid biosynthesis that underpins our interest in the enzyme.

The exact details of the quaternary structure of the *T. gondii* serine palmitoyltransferase (TgSPT) are currently unclear. We have identified two proteins likely

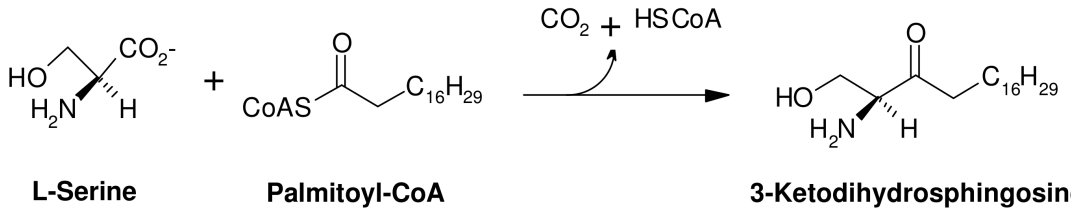
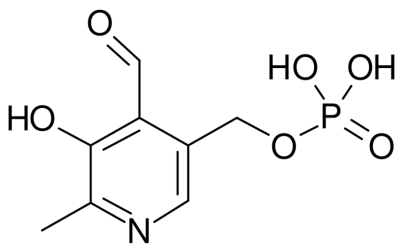


Figure 4: Serine plamitoyltransferase catalyses the Claisen condensation of L-serine with palmitoyl-CoA to form ketodihydrosphingosine.

to be involved in the condensation reaction shown in figure 4, though current analyses suggest that only one, the subject of the current study, contains all of the necessary active site residues, and is therefore likely to be fully active [37, 38]. It is hoped that more detailed structural information will further elucidate this matter. In yeast the analogous 'long chain base synthase' enzymes have a very complicated topology with a number of transmembrane domains and at least three different enzymes having been implicated in the condensation reaction [39, 40]. In the bacterium *S. paucimobilis* the enzyme is active as a water-soluble homodimer [41], the crystal structure of which has been solved [36] (see section 2.3.3). Ikushiro has also identified further bacterial analogues (from *Sphingobacterium multivorum*, *Sphingobacterium spiritivorum*, and *Bdellovibrio stolpii*) that have also been shown to exist as water-soluble homodimers [42]. Other work has gone into the identification and characterisation of the mammalian analogue to these enzymes, which is thought to exist as a heterodimer, though the exact functional relationship and a lack of structural information makes comparison difficult [43, 44].

2.3.1 Pyridoxal-5'-phosphate Binding

A significant number of enzymes use internally bound PLP (shown in 5) as a co-factor in their activity [45]; 4% of enzyme-catalysed reactions catalogued by the Enzyme Commission are PLP dependent [46]. PLP is a phosphorylated derivative of vitamin B_6 [47, 48], the main function of which seems to be the stabilisation



I

Figure 5: An active vitamer of B_6: Pyridoxal-5'-phosphate.

of anions created during reactions by the delocalisation of the negative charge in the cofactor's π system. The mode of binding is as an internal aldimine with the nitrogen atom in the side chain of a particular lysine residue in the structure [30, 49]. In the process of the reaction this internal aldimine undergoes a transaldimination to form an external aldimine with the amino acid reactant, this reactant then remains bound to the PLP through the course of the reaction until a final protonation step allows its release (as part of the product, KDS) and the reformation of the internal aldimine [30, 36].

Interestingly, phylogenetic analysis of sequences from eukaryotic, archebacterial, and eubacterial species has shown that a number of different families of PLP dependent enzymes have distinct evolutionary routes but have converged to have similar characteristics due to chemical necessities [50]. One difficulty encountered in designing drugs which target PLP dependent enzymes is due to the prevalence of PLP dependent enzymes. This has lead to a relatively small number of current drug targets and approved drugs for such enzymes. However, on-going research is looking at how this may be overcome, increasing specificity and reducing side effects [51].

TgSPT itself shows all the typical binding motifs expected for a PLP binding site, including the heavily conserved three residue His, Ala, Ser motif starting at

residue 291 which forms a significant part of the binding pocket. Other residues in the binding pocket are also seen to be present, such as Ser340, though these are generally less well conserved across the AOS superfamily [33, 36]. However, although it is likely that the overall fold is conserved in this family of enzymes, the specific details are in co-factor binding, and protein-ligand interactions are crucial to exploit structural information for future drug developments.

2.3.2 Known Inhibitors

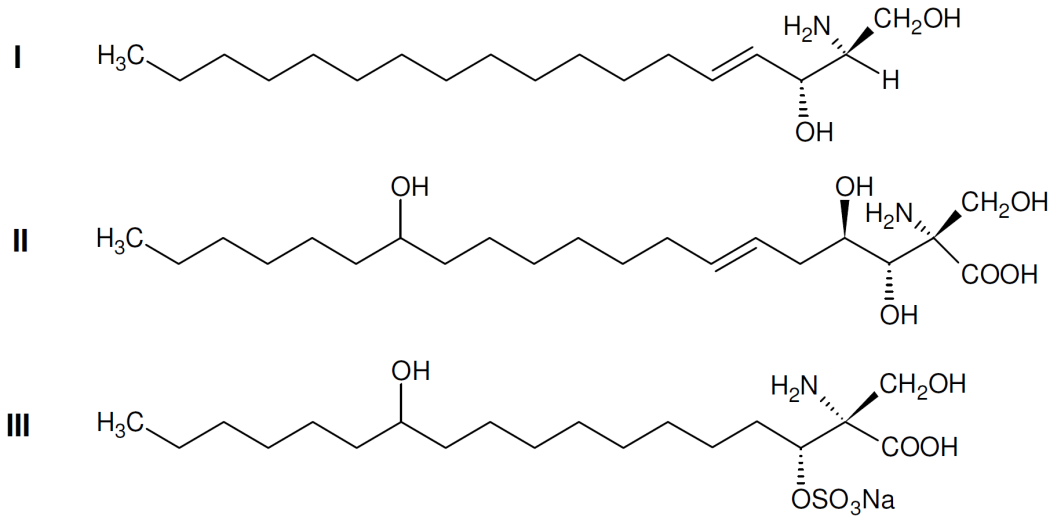


Figure 6: (**I**) Sphingosine; (**II**) Myriocin; and (**III**) Sulfamisterin.

Another factor which makes SPT a good drug target is the number of known inhibitors, both synthetic and natural, which make good potential starting points for drug development. Three such inhibitors are shown in figure 6. Two synthetic inhibitors, β-chloroalanine and L-cycloserine are also able to inhibit SPT, though their action is quite general, and they inhibit a wide range of PLP-dependent enzymes [52, 53]. The most widely used inhibitor is myriocin (ISP-1, thermozymocidin), which comes from the fungus *Isalia sinclairii* [54]. Its IC_{50} values for inhibition in mammalian cells have been shown to be in the nanomolar range.

For yeast however, the inhibition is much less strong, being in the micromolar range [55]. The reasons for this are not fully understood, though it is perhaps down to the action of the Sli1p protein [56].

Yamaji-Hasegawa has reported sulfamisterin, derived from the fungus *Pyncinidiella*, which inhibits both mammalian and yeast SPTs, with similar IC_{50} value to those of myriocin. They have also reported a number of synthetic derivatives which show varying degrees of inhibition, and give a good starting point for studying the specific mode of action of this family of inhibitors [57].

2.3.3 Serine Palmitoyltransferase from *Sphingomonas paucimobilis*

As noted above, the structure of the homologous SPT from *S. paucimobilis* has been solved to a resolution of 1.3Å. This has a 24% sequence homology with the conserved domain of TgSPT, and is the closest known structure. Figure 7 shows the homodimer of SpSPT with the PLP internal aldimine shown at the centre of both subunits. The uses of this analogue are two-fold; firstly it may be of use in phasing by molecular replacement, though the sequence identity is perhaps a little low, and it also provides a useful control for any functionality or substrate binding studies. This structure is, however, not suitable as a starting point for drug design because the homology is far too low to construct a reliable homology model based on it. In addition, this protein is from a prokaryotic organism, hence is very likely to be different in key ways from both the parasite as well as the human.

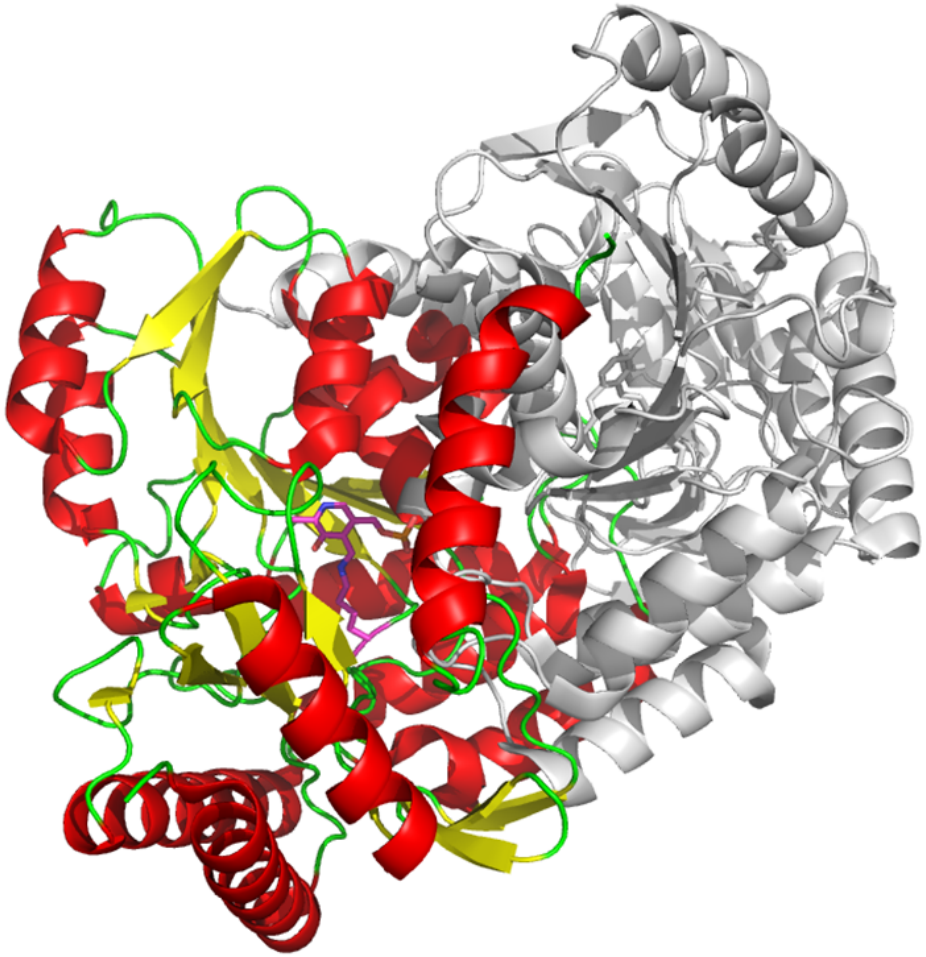


Figure 7: The ∼90kDa homodimer of serine palmitoyltransferase from *Sphingomonas paucimobilis* [36]. One monomer is shown coloured, and the other grey. In the coloured monomer α-helix is shown in red; β-sheet in yellow; and random coil in green. The pyridoxal-5'-phosphate (PLP) cofactor is shown in the centre of both monomers. (PDB Code: 2JG2)

3 Supramolecular Protein Assembly

3.1 Heat Shock Proteins

Stress responses in organisms were first noted in 1962 by Ritossa [58] who noted what he referred to as 'puffing' in the salivary glands of *Drosophila busckii*. A number of different environmental stimuli are known to cause stress responses such as heating, desiccation and chemical stress. The proteins associated with these responses are almost all known as heat shock proteins (HSP) and are ubiquitous in nature. They are classified further into families according to their molecular weight (e.g. proteins in the HSP100 will have molecular weights of ∼100kDA). Their mode of action is to act as molecular chaperones, helping the refolding of denatured or newly synthesized proteins, or preventing protein aggregation [59, 60].

3.1.1 Small Heat Shock Proteins

Small heat shock proteins are ubiquitous in nature and have been found in all three kingdoms of life. In the cell, they are known to associate with the nucleus, membranes and the cytoskeleton. The structure of their monomers can be broken down into three parts: The N-terminal extension; the C-terminal extension and a conserved 'α-crystallin' domain [61]. Whilst the conserved region acts as a basic building block, it appears that the N-terminal domain acts to aid oligomerisation [61, 62] and influences sub-unit dymanics and the C terminus determines the state of the oligomer and influences its stability and chaperoning function [61, 63, 64, 65].

3.1.2 Relationship to Human Disease: αB-Crystallin

αB-Crystallin is a heat shock protein and molecular chaperone found both in high concentrations in the eye [66], as well as at lower levels throughout the rest of the body, where its expression has been shown to be related to environmental stress [67, 68, 69]. It serves to provide the majority of the refractive power in the eye as well as acting to prevent other proteins from aggregating upon heat-shock and desiccation, keeping the lens clear. A number of mutations in this protein have been related to the formation of cataracts and myopathy (R120G [70], 450delA [71], 464delCT and Q151X [72]). Recent work has suggested whilst these mutations do not prevent the function of the protein, they do increase its tendency to self-associate and aggregate [73].

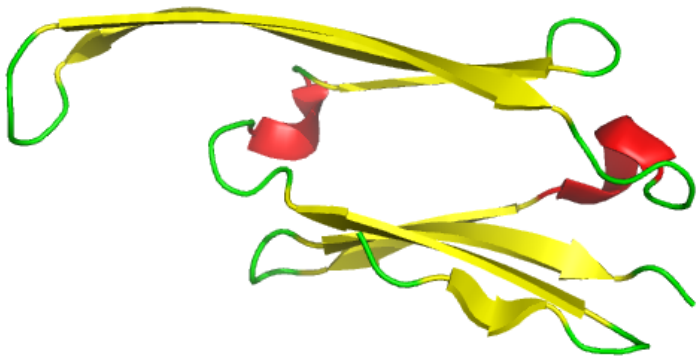


Figure 8: The truncated structure of human α-B-crystallin published by Bagnéris *et al.* [74] (PDB code: 2WJ7)

The structure of a truncated version of human α-B-crystallin has been published by Bagnéris *et al.* [74] (PDB code: 2WJ7), and is shown in figure 8. The structure has both N- and C-terminal mutations (detailed in appendix E), and therefore only shows the conserved α-crystallin domain as a dimer. Other work has used

cryo-electron microscopy to study the quaternary structure in greater detail [75, 76, 77]. Due, however, to its ready aggregation from solution during and after purification, it has not been possible to undertake crystallographic studies on the full-length protein to date, and it is therefore necessary to search out more stable analogues to get a fuller picture of the tertiary and quaternary structures. Furthermore, a number of studies have suggested that α-B-crystallin exists in multiple quaternary structures, which would make crystallisation impossible, as it relies on monodispersity.

3.2 Heat Shock Protein 16.5 from *Methanococcus jannaschii*

3.2.1 *Methanococcus jannaschii*

Methanococcus jannaschii is a hyperthermophilic methanogenic organism which has been found at deep sea vents [78]. Cultivation of the organism itself proves rather challenging, as they must be grown at 78℃ and 100kPa, however proteins from such organisms are generally found to be more robust and are therefore well suited to overexpression and purification as well as being more stable for crystallisation experiments. Its genome was also sequenced relatively early on, and it was therefore a significant target of the first phase of structural genomics approaches [79, 80].

3.2.2 Small Heat Shock Protein 16.5 as a Model Protein

A small HSP exists in *M. jannaschii* (MjHSP16.5)which has homology to α-B-crystallin. MjHSP16.5 was crystallised and the structure solved to a resolution of 2.9Å by Kim *et al.* in 1998 [81, 82, 83]. Since then the protein has been the subject of detailed study, though the full mechanism of its operation is still not fully understood [84, 85, 86]. Similar deletions to those which cause myopathy and

cataracts in αB-crystallin have been studied and shown to cause similar behaviour, though no crystallographic studies have been published to date.

Whilst MjHSP16.5 is only a small protein (147 amino acids, 16452gmol^{-1}) its functional unit is an octahedron made up of 12 dimers (Figure 30). One of the 3-fold axes in the octahedron was found to lie on a crystallographic axis, meaning that the asymmetric unit contained 8 independent monomers. Self-rotation functions showed that these were further connected by a number of non-crystallographic symmetry (NCS) axes allowing the use of restraints, simplifying the refinements and increasing the data:parameter ratio. In the original structure determination the first 32 amino acids were found to be disordered within the core of the octahedron, though it has been postulated that these play an important role in the initial folding of the icositetramer [82, 86].

A number of authors have used MjHSP16.5 as a model system, as well as looking into its general properties as a heat shock protein. Studies have considered the affects on the packing interfaces and their effect on oligomerisation [87], factors affecting its polydispersity in solution [88] and the factors affecting subunit exchange [89]. Through this work, we are now in a strong position to understand the behaviour of MjHSP16.5. This combined with the ease of its production from *E. coli*, and its stability in solution make it an ideal system for model study of the behaviour of heat shock proteins.

3.2.3 HSP16.5 and αB-Crystallin

There is a 17% sequence identity between MjHSP16.5 and human αB-crystallin and 31% sequence similarity. The structure of a truncated version of human α-B-crystallin has been published by Bagnéris *et al.* [74] (PDB code: 2WJ7). Whilst this structure only shows the protein as a monomer, the broad structural similarities between it and MjHSP16.5 are quite clear, as shown in figure 9. Due

to its instability and polydispersity in solution, it has however not been possible to undertake crystallographic studies on the full-length protein to date, and it is therefore necessary to search out more stable analogues to get a fuller picture of the tertiary and quaternary structures.

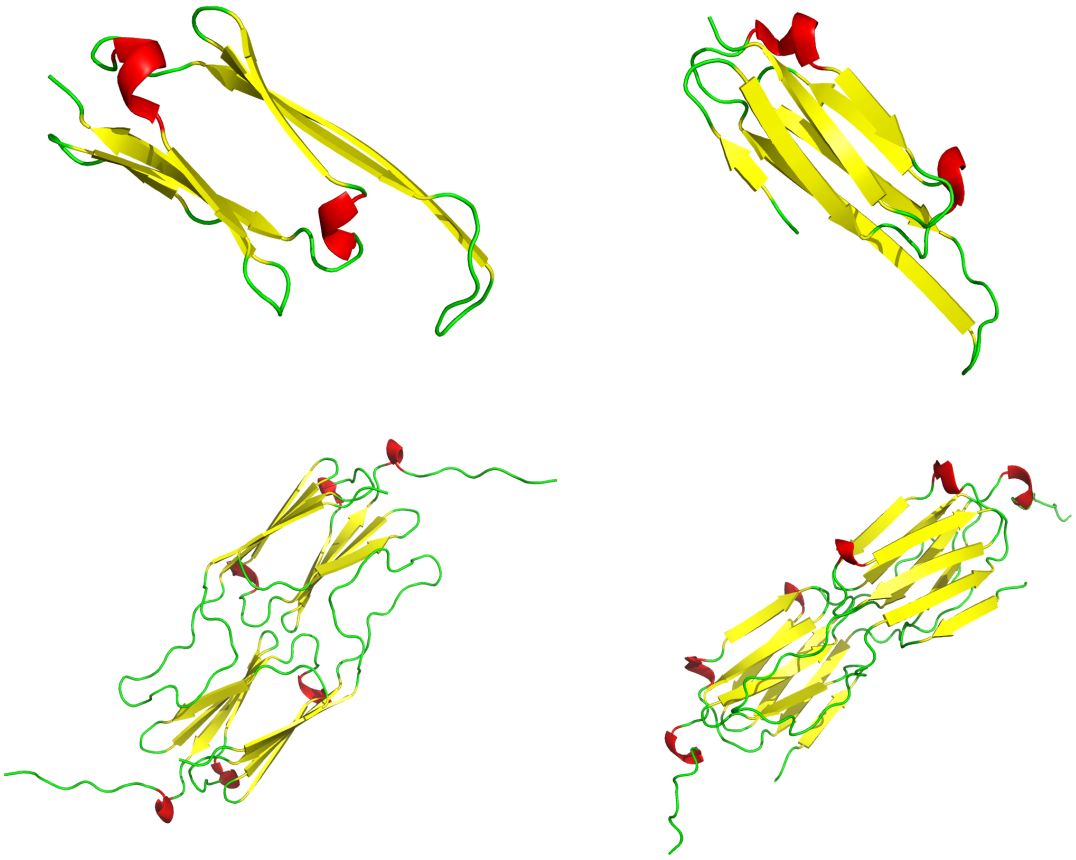


Figure 9: Similarities between structures: the top two images show a monomer of human α-B-Crystallin as solved by Bagnéris *et al.* [74] (PDB code: 2WJ7); the bottom two images show a dimer of MjHSP16.5. Whilst heavily truncated, the similarity in form is quite clear between the two. α-helix is shown in red; β-sheet in yellow; and random coil in green.

4 Experimental Background

Over the years a wealth of literature has been published on the various aspects of protein crystallography. For a broad, yet detailed introduction to the subject Bernhard Rupp's recent book on the subject proves a valuable starting point [90].

4.1 Protein Overproduction & Purification

The general thesis of protein overproduction is based upon our ability to 'hijack' the natural machinery of a cell for our own purposes. This is generally undertaken in either bacterial, insect or mammalian cells by the insertion of a plasmid, a section of circularised DNA, which contains information coding for the gene of interest as well as a selection of other genes, such as one coding for resistance to a particular antibiotic or genes coding for a particular system to control the expression of the gene of interest. Further to this, fusions such as a histidine tag may be included to aid solubility, reduce toxicity, or allow ready purification.

4.1.1 Plasmid Selection

A wide variety of plasmids are available for a number of different expression systems, allowing the selection of different fusion proteins and tags to aid overexpression and purification, a variety of which are shown in table 1. Most plasmids used here are from the pET family, which are used for expression in *E. coli* based on the T7/lactose system [91]. This is a powerful system which produces the protein, LacI, which binds to DNA in a particular sequence keeping expression to a minimal level by physically blocking the action of the DNA polymerase until the addition of D-galactose (or an analog isopropyl-β-D-thiogalactopyranoside (IPTG)), which inhibits the ability of LacI to bind the DNA, therefore releasing the DNA and allowing the polymerase to initiate expression. Plasmids also con-

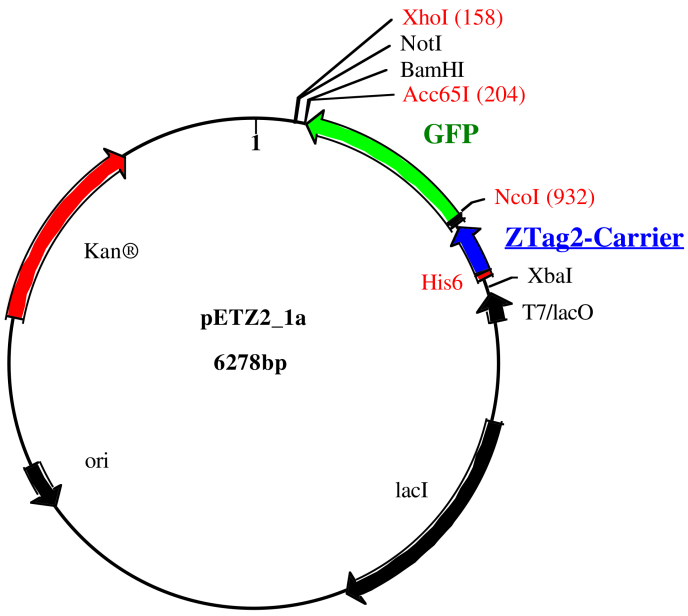


Figure 10: Schematic plasmid map of pETz_t21a, showing a number of features of the plasmid, including: kanamycin resistance, the ZTag2-Carrier fusion protein, the six histidine purification tag and lactose/T7 expression control system.[4]

tain a 'multicloning site'. This is a region engineered to contain a wide variety of different restriction enzyme recognition sites, allowing for easy insertion of coding sequences with complimentary ends.

4.1.2 Construct Design

A number of considerations are taken into account in designing a gene construct coding for a protein, or part of a protein, for crystallisation purposes. The two most important factors are functionality and solubility, as there is little biological worth in the structure of a non-functional protein, and crystals cannot be grown if the protein is not soluble. The general aim (where the full length protein is unsuitable) is to find a minimal, structurally homogenous, functional domain, which can then be expressed, purified and crystallised. A number of bioinformatics tools exist to aid in this process such as BLAST [95, 96], which allows ready

[4]Image courtesy of Gunter Stier, European Molecular Biology Laboratory, Heidelberg.

Tag	Properties
His_6	A tag comprising a repeat of six histidines, used for metal-affinity purification
Strep	A tag of eight amino acids (WRHPQFGG) which binds to streptavidin, and can therefore also be used for purification [92]
MBP	Maltose Binding Protein – Large fusion protein, which may improve expression levels, and give proper folding [93]
GST	Glutathione S-Transferase – Another large fusion protein, which may also improve expression levels, and give proper folding [94]

Table 1: Various commonly used tags and fusion proteins and their properties.

comparison of the protein sequences to others, showing any conserved domains of the protein, which in general correspond to the functional regions of a protein. Other computational tools are available to predict the secondary structure (e.g. Jpred 3 based at the University of Dundee [97, 98]) or find other features such as transmembrane helices (e.g. HMMTOP based at the Hungarian Academy of Sciences [99, 100, 101]), giving further information to aid design.

4.1.3 Cloning

Once the desired region has been identified, there are a variety of different strategies available to engineer a system through which the protein of interest can be produced:

Conventional Directional Cloning uses a combination of restriction and ligation to create a plasmid that may then be used to overexpress this protein. In this method, forward and reverse primers are designed, which may be used to amplify the desired region of the cDNA by the polymerase chain reaction (PCR) [102], while also engineering restriction sites onto the ends of the sequence to allow insertion of the construct into the desired plasmid. A variety of different restriction

enzymes are available, which cut DNA at a number of different recognition sites. Cut plasmid and insert with complimentary ends can then be ligated using a T4 DNA ligase and thus reforming the plasmid, allowing it to be inserted into the cell type of choice.

Topoisomerase Cloning This method exploits the ability of the enzyme DNA topoisomerase to act both as a restriction and ligation enzyme. This removes the need for the restriction step; instead recognition sites for the relevant topoisomerase are included in the primers used to amplify the protein of interest, which is then inserted into the plasmid directly from PCR. Furthermore, commercially available kits contain linearised plasmid with topoisomerase covalently bound to each 3' phosphate of the DNA, greatly increasing the speed and efficiency of the reaction, which in practice only takes around five minutes.

Ligation Independent Cloning This method negates the need for both restriction and ligation by exploiting the 3' to 5' exonuclease activity of T4 DNA polymerase in the presence of a single deoxyneucleotide (dNTP) [103]. To do this, both the insert and the plasmid contain a specific twelve nucleotide end sequence, lacking one pair of nucleotides. Then, in the presence of the one of the missing dNPTs (one for the insert and the other one for the plasmid) the T4 DNA ligase cleaves all the 3' nucleotides up to the first concurrence of the nucleotides missing from the end sequence, leaving complementary single stranded DNA. These overhangs are then long enough to negate the requirement for ligation, being long enough to anneal independently, and the backbone is then rejoined by the host DNA repair and expression machinery.

4.1.4 Overexpression

E. coli strain BL21(DE3) are most commonly used for protein expression. A number of different variants of the strain are available such as BL21 AI, which contains an additional layer of expression control to aid in the expression of toxic proteins. Other variants contain plasmids that supplement the expression machinery, by introducing rare codons not normally found in *E. coli*, or coexpressing lysozyme, to help with cell lysis and extracting the protein of interest.

In standard expression, as the plasmid containing the DNA for the protein of interest also contains the genes encoding for LacI (see section 4.1.1 for more details), no protein is expressed until IPTG is added to the growth medium. Cultures are first grown whilst monitoring their optical density (OD) at 600nm, giving an indication of the stage of the cells' growth. Expression is then induced with IPTG at between OD 0.3 and 0.6, when the cultures are in their exponential phase, and growing most rapidly. They are then grown for 3hrs, generally at reduced temperature, after which, cells are harvested by centrifugation.

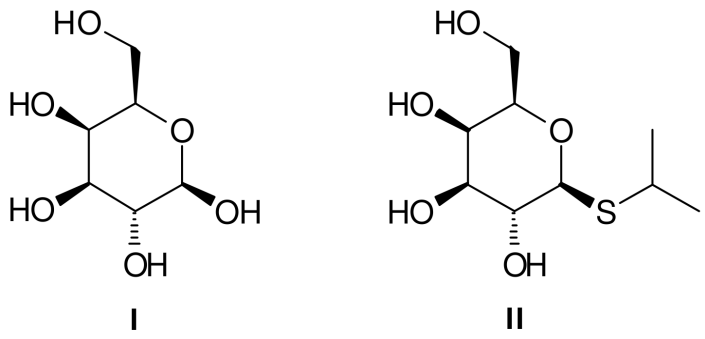


Figure 11: (**I**) β-D-Galactopyranoside (D-Galactose) (**II**) Isopropyl-β-D-thiogalactopyranoside (IPGT).

Autoinduction For some proteins, standard expression protocols, even with reduced levels of IPTG, may prove to harsh and can cause problems, particularly

with solubility. A more gentle way to bring about expression is to use autoinduction as developed by Studier [104]. This technique relies on the natural response of *E. coli* to conditions where there are multiple sources of energy available. In the case of autoinduction, we use glucose, lactose and glycerol. In the first place, the cells use the glucose as their food source, and the cell uptake and metabolism systems relating to lactose metabolism are silenced. However, after a period of time, the rapidly growing cells use up all the glucose, and the *lac* operator is de-repressed. Consequently, the gene encoding the T7 RNA polymerase in our plasmid is also de-repressed, and our protein begins to be expressed as well.

4.1.5 Purification

The most straight forward route to purification is to design all constructs with a multi-histidine fusion tag to allow the use of the quick and robust procedure of nickel affinity chromatography. This tag can then be connected to the protein by a linker which contains the recognition sequences for a site-specific protease of which there are a number. Perhaps the most commonly used of these is the 27kDa NIa protease from the tobacco etch virus (TEV) [105, 106], which recognises the linear epitopes of the form `EXXYXQG` or `EXXYXQS`, cutting between the Q and the G or the Q and the S. Another commonly used protease is the 3C protease from human rhinovirus, which recognises sequences of the form `LEVLFQP` or `LEVLFGP` [107]. Furthermore, as in both cases the protease is engineered with a His_6 tag, the cleavage reaction mixture can once again be passed over a nickel affinity column, and thereby all protease, cleaved tag, uncleaved fusion protein and other impurities with nickel binding sites is removed from the mixture, leaving the cleaved protein of interest pure.

4.2 Crystallisation

Crystallisation occurs when a protein becomes oversaturated in solution making nucleation a more favourable process. Due to the instability of proteins in solution, this is a difficult process to undertake with aggregation being a significant difficulty. As such it is necessary to find a condition in which the protein is more stable as its concentration increases. A number of different techniques are used to promote such a situation and a detailed handling of these can be found in Terese Bergfors' book on the subject [108]. A general scheme would involve an initial high-throughput screen using a crystallisation robot [109] followed by rounds of optimisation of positive conditions to yield better crystals. Whilst visual inspection may give an immediate feedback of the suitability of conditions, in-house testing of crystals grown is used to identify the best conditions for X-ray diffraction.

The most common technique used for protein crystallisation is vapour diffusion. Small (0.05-10μl) droplets are made containing similar volumes of a concentrated protein solution and of a crystallisation solution which is usually made up of a buffer, a precipitant and often some kind of additive such as a salt. These are then sealed along with a much larger reservoir of the crystallisation solution. Over time the system equilibrates, changing the concentration in the drop and, in the preferred scenario, promoting crystallisation.

The use of such high-throughput screening methods has vastly improved the speed and possibility of protein structure determination. Whereas previously the amount of time available has constrained the ability to tackle more challenging structures, it is now entirely possible to set-up over 1000 crystallisation experiments in an afternoon's work, removing a significant bottleneck to progress. Indeed, the design of such screens has become a science in itself, bringing in to play such ideas as genetic algorithms and factorial design increasing success rates still

further.

4.2.1 Cryoprotection

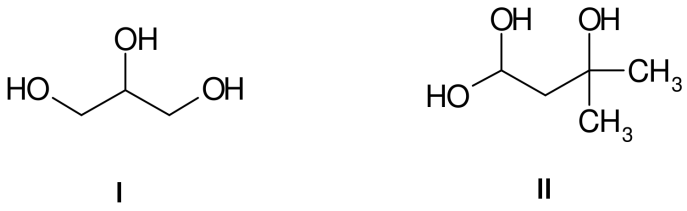


Figure 12: Two cryoprotectants: (**I**) Glycerol and (**II**) 2-methyl-2,4-pentanediol (MPD).

In order to reduce dynamic disorder and slow down radiation damage (by decreasing the mobility of radical species created), data are collected at 100K [110]. This can lead to the formation of ice within crystals, reducing data quality, and potentially destroying the crystal. To prevent this a selection of different 'cryoprotectants' are used, such as glycerol and 2-methyl-2,4-pentanediol (MPD) (Figure 12). These may either be included in the crystallisation solution as the precipitant or where this is not possible the crystal may be incubated in a suitable solution before freezing [111].

4.3 X-Ray Diffraction

4.3.1 The Ideal Crystal

An idealised crystal consists of regular arrays of atoms and molecules which repeat infinitely throughout the solid. As such, one can describe the crystal by its 'unit cell'; the most simple repeat unit within a crystal, being defined by six parameters; three angles (α, β and γ), and three dimensions (a, b and c), such that α lies between b and c, opposite a *etc.*. Further to this, one can define planes within

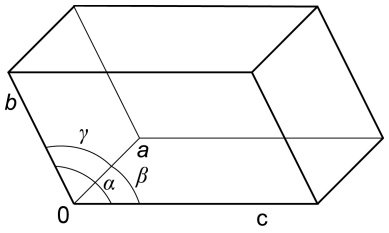


Figure 13: A schematic representation of a unit cell.

the crystal which repeat infinitely. These are defined by the Miller indices h, k and l. A plane h, k, l intercepts the unit cell axes at points a/h, b/k and c/l. By convention, negative indices are written as $\bar{n}$ for $-n$. Indices are also expressed such that the greatest common divisor is 1.

4.3.2 Bragg Diffraction

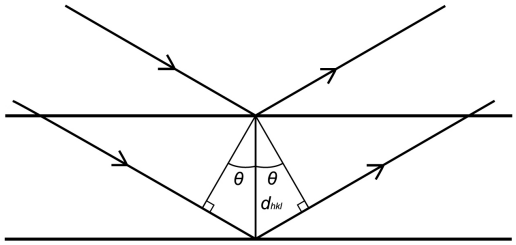


Figure 14: A schematic representation of Bragg diffraction.

As observed by von Laue, a beam of X-rays may be diffracted when passing through a crystal. This occurs due to the interaction between the X-ray and the electrons in the solid. Bragg peaks occur when the X-rays are scattered in phase (when an integer number of waves fit between two Miller planes), as any other interference would interact destructively. The net effect of this gives rise to the discrete pattern of spots observed on a diffractogram. The relationship between the wavelength of X-rays used (λ), the spacing between planes (d_{hkl}) and the angle

between the plane and the diffracted beam (θ) is defined by the Bragg equation:

$$\lambda = 2d_{hkl} \sin\theta \tag{1}$$

4.3.3 Space Groups

We have already considered a representation of the translational symmetry in a crystal in terms of three angles (α, β and γ) and three dimensions (a, b and c). While in theory an infinite number of different unit cells might be defined for a particular crystal a number of conventions exist which allow a standardised representation. For instance, in a system where $a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$ there is no metric symmetry within the unit cell and the system is said to be triclinic. At the opposite extreme, a crystal where and $a = b = c$ and $\alpha = \beta = \gamma$ must contain four three-fold rotation axes and is said to be cubic. In all seven crystal systems are defined. These are further subdivided into fourteen 'Bravais lattices' on a basis of the position of lattice points within the cell. Lattices can be either primitive P, where lattice points only exist at the corners of the cell, body centered I, face centered F, where an extra lattice point exists on each face of the cell or centred on a single face A, B or C, though it is important to note that not all combinations are unique. These are then further subdivided on a basis of internal symmetry operations, such as mirror planes and rotation axes, into a total of 230 space groups, which are listed in [112].

This picture can be simplified somewhat for protein crystallography. Proteins are enantiomerically pure, being composed entirely from L-amino acids, a mirror plane or an inversion operation would transform the protein to the D- form which does not occur naturally. As such, protein structures are restricted to non-mirror and non-inversion symmetry operations. This reduces the number of space groups available to the 65 chiral space groups.

4.3.4 The Structure Factor

In the diffraction experiment we measure the intensity, I_{hkl}, of a set of reflections. Certain corrections are applied to these to give a concept, the structure factor F_{hkl}:

$$I_{hkl} = |F_{hkl}|^2 \cdot LP \cdot A \tag{2}$$

Where LP represents a polarisation and geometry correction and A an absorption correction. In turn, in a simplified picture of a crystal, F_{hkl} may be related to the contents of the unit cell by the Fourier transform:

$$\begin{aligned} F_{hkl} &= \sum_{j}^{atoms} f_j(\theta) \exp[2\pi i(hx_j + ky_j + lz_j)] && (3) \\ &= \sum_{j}^{atoms} f_j(\theta) \cos[2\pi(hx_j + ky_j + lz_j)] + \\ &\quad \sum_{j}^{atoms} f_j(\theta) \sin i[2\pi(hx_j + ky_j + lz_j)] && (4) \end{aligned}$$

Where $f_j(\theta)$ is a function relating the scattering power of each atom in relation to the scattering angle 2θ (the angle between the incident and diffracted beam) and x_j, y_j and z_j are the fractional coordinates of each atom in the unit cell.

4.3.5 Systematic Absences

Depending on the particular symmetry within a crystal its diffraction pattern may display certain systematic absences; expected Bragg peaks which are not observed. These are where geometry within a crystal causes certain reflections to have an intensity of zero. Well established rules can tell us where we might see systematic absences. For instance, in the body centered cubic system, for a

monatomic crystal equation 3 reduces to the following two cases:

$$F_{hkl} = \begin{cases} 2f_j(\theta), & h+k+l \ \text{ even} \\ 0, & h+k+l \ \text{ odd} \end{cases} \tag{5}$$

Many other crystallographic features, such as screw axes, glide planes and other forms of centering also have systematic absences associated with them. The combination of this information with the unit cell parameters determined can be used to help identify the particular space group of the crystal's unit cell.

4.3.6 Electron Density

The exact relationship between the electron density and the diffraction pattern can be described by a Fourier transform:

$$F_{hkl} = \int_{cell} \rho(xyz) \cdot \exp[2\pi i(hx + ky + lz)]dV \tag{6}$$

with $\rho(xyz)$ being a continuous positive function representing the electron density in terms of the fractional coordinates x, y and z. This represents a very complicated function to compute, and is simplified by treating the structure as being composed of discrete spherical quantities as in equation 3. This reduces equation 6 to the summation:

$$F_{hkl} = \sum_j f_j(\theta) \cdot \exp(-8\pi^2 U_j \sin^2\theta/\lambda^2) \cdot \exp[2\pi i(hx_j + ky_j + lz_j)] \tag{7}$$

where $f_j(\theta)$ is the atomic scattering factor, and U_j a displacement parameter, representing the mean squared displacement of a particular atom. In cases where sufficient data are available this may be further improved by the use of an anisotropic displacement matrix, allowing for different displacement in three directions, mod-

elling each atom's displacement with an ellipsoidal probability distribution.

Equation 6 can also be restated in terms of $\rho(xyz)$, giving us an expression for the electron density which is the physical quantity we aim to measure, and from which we infer the position of the individual amino acids. This is the reverse Fourier transform of the diffraction pattern, however unlike equation 6, because the diffraction pattern is made up of discrete reflections, rather than a diffuse pattern, the Fourier transform is immediately a summation:

$$\rho(xyz) = \frac{1}{V} \sum_{h,k,l} F_{hkl} \cdot \exp[2\pi i(hx + ky + lz)] \tag{8}$$

4.3.7 The Phase Problem

Every diffracted wave has both an amplitude and a phase. In the diffraction experiment we are only able to measure the amplitudes. Likewise, the structure factors may be considered as vectors for which only the magnitude is known, in the absence of any phase information. An analogy to this situation might be to consider being given a copy of a poem without indication of in what order the words are to be placed; any number of combinations are possible, but only certain will make any sense, and with luck one might make significantly more sense than any other. For this reason we must use some prior knowledge to aid our solution.

Mathematically this is seen by restating equation 8 with explicit statement of the phase information contained within F_{hkl}:

$$\rho(xyz) = \frac{1}{V} \sum_{h,k,l} |F_{hkl}| \cdot \exp[i\phi(hkl)] \cdot \exp[2\pi i(hx + ky + lz)] \tag{9}$$

4.3.8 The Patterson Method

In a Patterson synthesis, we replace the corrected observed amplitudes $|F_o|$ with their squares. This effectively sets all phases to zero. The consequences of this are

seen in the following equation which may be derived from equation 9, by making this change:

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)] \tag{10}$$

The result of this new Fourier synthesis is no longer the electron density, but instead a function $P(uvw)$. This holds many similarities to an electron density distribution; its coordinates are still fractions of the unit cell parameters and it is a continuous function repeated in each unit cell. However, its peaks no longer represent the positions of individual atoms, instead they represent vectors between pairs of atoms in the structure, their heights being proportional to the product of the atomic numbers of the two atoms. Due to a convolution effect the peaks in the Patterson function are twice as wide as those in the corresponding electron density map leading to an often indistinguishable mass of peaks.

Where heavy atoms are present in the structure, their contribution to the Patterson function is more pronounced than those of lighter atoms. This can give us information about the positions of these atoms allowing the calculation of basic phases. From these it is often possible to discern further features in the electron density map leading to cycles of map improvement followed by phase improvement, giving a full structure solution.

4.3.9 Molecular Replacement

Another technique for finding atomic positions in the unit cell from a Patterson synthesis is the use of a known molecular fragment to calculate a set of peaks which can then be found within the Patterson map, leading to phase information. This is of particular use in protein crystallography, where the technique is known as molecular replacement, as evolution has kindly lent us a sets of similar proteins.

Where the structure is already known for a protein with a sequence identity of ∼30% or above, this structure might be used to make rotation and translation searches of the Patterson function again giving phase information [113, 114, 115].

4.3.10 Single/Multiple Isomorphous Replacement

A difficulty encountered in the use of the Patterson method in protein crystallography is the general lack of heavy atoms. In some cases, for instance haemoglobin, these are already present and allow solution more readily, but often it is necessary to introduce them to a material. Indeed, even in the case of haemoglobin it may prove difficult to identify the location of the haem group due to the very large number of other atoms present. Two general strategies exist to pursue this. The first involves the initial growth or later 'soaking' of crystals in a solution containing a particular heavy atom additive which might then bind preferentially to a particular site in the crystal. The second involves the introduction of the heavy atom directly into the protein by the substitution of the sulphur containing amino acid methionine with its selenium derivative.

In both cases the we are left with two datasets, one with the heavy atom and one without. If the two materials are isomorphous, it is possible to calculate the contribution of the heavy atoms to the diffraction and thus determine their positions quite accurately, allowing the calculation of phase information relating to the heavy atoms. Now consider the structure factors as vectors. An isomorphous replacement experiment with one derivative gives us the following equation:

$$F_{derivative} = F_{protein} + F_{heavy\ atom} \tag{11}$$

As we know the magnitude of each vector and the orientation of $F_{heavy\ atom}$ there are two possible solutions for the orientation of $F_{protein}$. If we further prepare a

second derivative, we will have a second equation allowing for the unambiguous solution of the phase information.

4.3.11 Single/Multiple Anomalous Diffraction

For light atoms the function $f_j(\theta)$ may be considered as constant with respect to the wavelength of the radiation used. For heavier atoms (those heavier than sulphur), this is not the case. Around their X-ray absorption edge, there is a distinct change in their scattering power with change in energy due to resonance between the X-ray waves and electronic transitions from bound atomic orbitals. This may be considered as:

$$f_j^{total}(\theta,\lambda) = f_j^0(\theta) + |f_j^{\Delta}(\lambda)|e^{i\delta} \quad (12)$$

$$= f_j^0(\theta) + f_j^{'}(\lambda) + if_j^{''}(\lambda) \quad (13)$$

Where $f_j^0(\theta)$ is the normal scattering factor, $f_j^{\Delta}(\lambda)$ the anomalous scattering factor and f' and f'' the real and imaginary components of the scattering factor. Two datasets collected either side of the absorption edge would therefore give different values of F_{hkl}, the differences relating to the anomalous diffraction, leading to a similar situation as above, giving further phase information [116]. The values of f' and f'' may be obtained from X-ray absorption spectra by the relationship:

$$f''(E) \propto E \cdot \mu_a(E) \quad (14)$$

Where μ_a is the atomic absorption coefficient and E the energy of the X-ray wave. Real values may then be obtained by Kramers-Kronig transformation from f'' values [117].

The most common element used for anomalous diffraction experiments is sele-

nium as outlined in Section 4.3.10. Indeed in seeking to combine as much information as is available, a general scheme would normally involve combinations of both isomorphous replacement and anomalous diffraction.

4.3.12 Data Reduction

With the advent of CCDs for data collection, crystallographers moved away from directly measuring the intensity of each reflection, and it is therefore necessary to calculate the intensity of each spot using a process know as data reduction. This first involves harvesting spots on frames and indexing them. Indexing is the process of assigning the Miller indices to each reflection, which first requires the determination of the unit cell of the material and the orientation matrix of the crystal (relating the laboratory axes (xyz) to the internal axes of the crystal (abc)). As each reflection is spread across a number of frames, it is then necessary to integrate and merge each of them. This process involves fitting a profile to each spot allowing the calculation of the total intensity across all frames for each reflection. Certain corrections are applied to the data to account for factors such as absorption in the crystal and the size of the crystal giving a final intensity for each reflection.

5 Materials & Methods

5.1 Bioinformatics

Gene databases were searched using the Basic Logical Alignment Search Tool (BLAST) web-server [95, 96], which shows regions of homology between the submitted sequence and other known proteins. Protein structures were searched using the RSCB PDB [118]. More detailed sequence alignments were performed using the EBI ClustalW web-server [119, 120, 121].

Basic secondary structure features were identified using the Jpred 3 web-server [97, 98], and transmembrane helices identified using the HMMTOP web-server [101, 100]. Properties of potential constructs such as mass, amino acid composition, theoretical pI, and theoretical stability were calculated using the ExPASy ProtParam web-server [122, 123].

Homology models were calculated using MODELLER [124]. Protein sequences were used in searches for similar sequences in the PDB [118], and the sequence with the highest similarity used to calculate the model. All models were analysed and least-square superpositions were performed with Coot [125].

5.2 Standard Cloning & Expression

Work performed at Durham followed a standard low-throughput DNA cloning and protein overproduction strategy. Plasmids were synthesised through PCR, restriction, and ligation. Expression was trialled on the decilitre scale, and large-scale expression undertaken on a litre scale.

5.2.1 DNA Synthesis & Cloning

Cloning was undertaken in the *E.coli* strain DH5α. Small quantities of plasmid were obtained and purified using the Qiagen Spin MiniPrep system with a Sigma

Table 2: Examples of DNA primers used for the cloning of truncated TgSPT constructs.

Name	**Sequence**
5*NcoI*TgSPT47	CATG **CCATGG** CC AACGACGAAGTGAGCTGG
5*NcoI*TgSPT181	CATG **CCATGG** CC CCTTTCGCCTACGAAGTCT
5*HindIII*TgSPT181	CCC **AAGCTT** GC CCTTTCGCCTACGAAGTCT
3*NotI*TgSPT_STOP	ATAAGAAT **GCGGCCGC** TTA TCGGAGCATGTCAGTGGGTGGG

1-14 microcentrifuge, and larger quantities using the Spin MidiPrep system with a Boeco U320 centrifuge. Linear DNA was purified by the QIAquick PCR Product Purification system with a Sigma 1-14 microcentrifuge. FideliTaq from the USB Corporation was used for proof-reading PCR and Promega GoTaq for more general screening purposes. All primers were ordered from Sigma-Aldrich. T4 DNA ligase and all restriction enzymes were from Promega. Throughout the BioRad Sub-cell GT agarose gel electrophoresis system was used for analysing DNA samples. Ligation reactions were incubated for 16h at 4℃, and then used to transform SoloPack Gold cells from Strategene by heat shock at 42℃ for 90s following the manufacturer's standard protocol. Positive cell lines were then identified by colony PCR. The presence of correctly ligated insert was further verified by restriction digestion analysis and DNA sequencing.

5.2.2 Protein Overproduction & Purification

Various expression strains were trialled for protein expression, and are tabulated in table 3. Chemically competent cells were transformed with plasmid after incubation on ice with the plasmid for 30 minutes by heat shock at 42℃ for 90 seconds, incubation on ice for 2 minutes, followed by the addition of 1250μl of SOC (Super Optimal Broth with Catabolite Repression: Tryptone (pancreatic digest of casein) 2% (w/v), Yeast Extract 0.5% (w/v), NaCl 8.6 mM, KCl 2.5 mM, MgSO4 20 mM, Glucose 20 mM), and incubation with shaking at 120 rpm at 37℃ for one

Expression Strain	Producer	Features
BL21(DE3)	Invitrogen	Standard T7 expression strain
BL21(DE3)pLysS	Promega	BL21(DE3) expression strain, but with an additional plasmid encoding for T7 lysozyme, to lower background expression levels of target genes
BL21-AI	Invitrogen	BL21(DE3) expression strain, but with the T7 RNA polymerase gene controlled by the *araBAD* promoter, preventing expression without the presence of arabinose, reducing background expression of target genes
BL21-CodonPlus	Stratagene	BL21(DE3) expression strain, but with additional copies of tRNAs which are rare in *E. coli*
Tuner	EMD	*lacZY* deletion mutant of BL21(DE3), allowing regulation of expression levels by varying the concentration of IPTG
Rosetta	EMD	Tuner expression strain, including additional copies of tRNAs which are rare in *E. coli*
Rosetta-pLysS	EMD	Rosetta expression strain, with an additional plasmid encoding for T7 lysozyme to lower background expression levels of target genes

Table 3: Details of various cell lines used for standard expression screening, and protein production.

hour, before spreading on LB (Luria Bertani Broth: Tryptone (pancreatic digest of casein) 10 g/L, Yeast Extract 5 g/L, NaCl 5 g/L) agar plates, and incubation overnight at 37℃. For large scale IPTG-induced expression, 1l flasks of LB, and the appropriate antibiotic(s) were then inoculated with 1ml of overnight culture and grown at 37℃ with shaking at 120 rpm, and their OD 600nm monitored. Cultures were induced when they reached an OD of 0.4 by the addition IPTG to a final concentration of 1mM and grown at various temperatures for 3 hours. Cells were harvested by centrifugation in a Beckman Coulter Avanti J-20XP centrifuge at 5000×g for 15 minutes.

Autoinduction Some large-scale cultures were also grown by autoinduction. The composition of media used was the same as was described by Studier in 2005 [104]. Inoculated 1l cultures were grown at 37℃ with shaking at 120 rpm, until their OD at 600nm reached 0.5, at which point their temperature was reduced to 20℃. Cells were harvested by centrifugation in a Beckman Coulter Avanti J-20XP centrifuge at 5000×g for 15 minutes.

Pellets were resupended in a lysis buffer containing EDTA-free protease inhibitor cocktail from Roche. Cell lysis was performed by 6 cycles of sonication with a Sanyo Soniprep 150 (30s on, 30s off) on ice. Cell debris was then harvested by centrifugation at 25,000×g for 20mins. Crude lysate was then filtered through 0.45μm and 0.22μm filters. Nickel affinity purification was performed using a GE Healthcare ÄKTA Explorer with 5ml HisTrap FF columns. Proteins were first bound to the column under low imidazole concentration (10mM) and then eluted over an increasing imidazole gradient (10–1000mM). Absorbance at 280nm was measured to identify fractions containing protein which were then pooled and dialysed overnight with remove imidazole. Where tag cleavage was required, the appropriate protease was included with the protein for dialysis, and a second nickel affinity purification was undertaken, uncleaved protein, fusion tag and protease remaining bound to the column with cleaved protein in the flow-through. Protein samples were then analysed by SDS PAGE, using the Bio-Rad Mini-PROTEAN Electrophoresis system.

5.3 High-throughput Expression Screening

High-throughput work was undertaken at the Oxford Protein Production Facility at the Harwell Science and Innovation Campus in Oxfordshire. The work is based a on 96-well plate format, generally producing 48 different constructs of varying length in purpose made vectors of the 'pOPIN' series. Cut vector and in-

Vector	Parent Vector	Features
pOPINE	pTriEx2	C-terminal His_6 tag.
pOPINF	pTriEx2	3C-protease-cleavable His_6 tag.
pOPINJ	pTriEx2	3C-protease-cleavable His_6 and GST tag.
pOPINS3C	pET28a	3C-protease-cleavable His_6 and SUMO tag.

Table 4: Details of the four vectors used for the high throughput expression screening protocols.

Vector	Extension	
pOPINE	Forward	`AGGAGATATACCATG`
	Reverse	`GTGATGGTGATGTTT`
pOPINF	Forward	`AAGTTCTGTTTCAGGGCCCG`
	Reverse	`ATGGTCTAGAAAGCTTTA`
pOPINJ	Forward	`AAGTTCTGTTTCAGGGCCCG`
	Reverse	`ATGGTCTAGAAAGCTTTA`
pOPINS3C	Forward	`GCGAACAGATCGGTGGT`
	Reverse	`ATGGTCTAGAAAGCTTTA`

Table 5: Details of the InFusion primer extensions for each of the four vectors used for the high throughput expression screening protocols.

sert are combined using the InFusion system of ligation-independent cloning from Clontech. Expression screening is then undertaken on a 2.5ml scale using IPTG and autoinduction in two different expression strains. Purification uses a Qiagen BioRobot 8000 with Ni-NTA magnetic beads.

5.3.1 DNA Synthesis & Cloning

Constructs were designed using the standard set of bioinformatics tools as described in section 5.1, and primers designed with specific extensions for each vector used. Details of the four vectors used are shown in tables 4 & 5.

Template Amplification by PCR PCR reactions were undertaken in a Veriti 96-well thermal cycler from Applied Biosystems. Initial reactions were undertaken using KOD Xtreme Hot Start DNA Polymerase Mastermix from Novagen, using

the manufacturer's standard protocol. A second round of reactions was undertaken using Phusion Flash Mastermix from New England Biolabs and used for conditions which had not amplified well with the KOD Xtreme. PCR products were then analysed by agarose gel electrophoresis using a 1.25% TBE/agarose gel using 4μl of Hyperladder I from BioLine as reference, and SYBRSafe Stain from Invitrogen for visualisation under UV light, and were loaded using 0.25% w/v bromophenol blue in 30% v/v glycerol as a loading buffer (5μl product to 2μl buffer). Following removal of the template DNA by treatment with DpnI, the amplified DNA was then purified using the Agencourt AMPure XP magnetic bead purification system from Beckman Coulter, using the manufacturer's standard protocol.

Ligation Independent Cloning InFusion reactions were performed using lyophilised enzyme and buffer from the manufacturer, with 100ng of template and between 10 and 250ng of purified insert in a 20μl reaction volume. Reactions were then held at 42℃ in the Veriti thermocycler. After 30 minutes the reactions were transferred onto ice, and diluted with 40μl TE buffer (10mM Tris pH 8.0, 1mM EDTA). 5μl aliquots of these reactions were then transformed immediately into OmniMax II chemically competent cells from Invitrogen, by 30 second heat-shock after 30 minutes incubation on ice. Cells were then returned to ice for 2 minutes before the addition of 300μl of Power Broth from Molecular Dimensions. These were then incubated without shaking for 1 hour at 37℃.

24-well agar plates were then prepared (2 wells for each reaction) with 1ml of LB agar, including the appropriate antibiotic, as well as X-gal and IPTG to allow for blue-and-white screening. Before plating out the cells, 20μl of Power Broth was added to each well, to allow for an even spread of colonies. Then for each reaction 5μl and 20μl of the concentrated cells at the bottom of the unshaken

transformation tubes were added to two separate wells. The plates were then shaken vigorously, and left to dry for 15 minutes, before incubation upside-down overnight at 37℃.

Verification of Colonies If the reactions have been successful, then blue colonies should only account for 10% of all colonies. If this is the case, then two colonies are picked for each construct, leading to a roughly 95% success rate. Due to the use of blue-and-white screening, it is generally felt safe to purify the plasmids before verification PCR.

Deep well blocks were prepared with 1.5ml of Power Broth, and the correct antibiotic. White colonies were then picked by touching them with 200μl pipette tips, and leaving them to stand in the Power Broth for a short while. The filled plates were then sealed with gas-permeable seals, and shaken at 200 rpm at 37℃ overnight. Glycerol stocks were produced and stored at -80℃ by addition of 100μl of the overnight culture to 100μl 30% v/v glycerol. The cells from the remaining overnight culture were then harvested by centrifugation for 15 minutes at 5000×g using a Beckman Coulter Avanti centrifuge with JS5.3 rotor, the supernatant then being decanted off.

Plasmids were then purified using the Promega Wizard SV 96 Plasmid DNA Purification system on the Qiagen BioRobot 8000, following the manufacturer's protocol, and verified again using KOD Xtreme Hot Start DNA Polymerase Mastermix from Novagen, and the same method of visualisation as previously.

5.3.2 Protein Overproduction & Purification

Transformation into Expression Strains Two different expression strains were trialled: B834(DE3) and Rosetta(DE3)LysS. Chloramphenicol must be used if the pRARE-LysS functionality is required to be kept. Transformation is under-

taken using the same procedure as previously, but with 3μl of purified plasmid being added to 50μl of competent cells, and only one well used per transformation, with 30μl of cells being plated out (it does not prove so vital to pick a single colony, as the purified plasmid was prepared from a single colony, though single colonies are still desirable).

Expression Testing with IPTG Induction Individual colonies were picked using 200μl pipette tips, and used to inoculate 0.5ml of Power Broth, supplemented with 1% w/v glucose and the appropriate antibiotic, in 96-well deep well plates. These were then sealed with gas-permeable seals, and grown overnight at 37℃ with shaking at 240 rpm.

A set of 24-well deep well plates were then prepared with 2.5ml of Power Broth, supplemented with 1% w/v glucose and the appropriate antibiotic. These were then inoculated with 62.5μl of the overnight culture, and grown at 37℃ with shaking at 240 rpm for 3-5 hours, until the OD at 595nm was roughly 0.5. 12.5μl of 100mM IPTG in water was then added to each well, to give a final concentration of 0.5mM, and the cultures were grown overnight at 20℃ with shaking at 200 rpm. The cultures were then transferred back to 96-well deep well blocks, and the cells were harvested by centrifugation for 15 minutes at 5000×g using a Beckman Coulter Avanti centrifuge with JS5.3 rotor, the supernatant then being decanted off.

Expression Testing with Autoinduction As with IPTG, individual colonies were picked using 200μl pipette tips, and used to inoculate 0.5ml of Power Broth, supplemented with 1% w/v glucose and the appropriate antibiotic, in 96-well deep well plates. These were then sealed with gas-permeable seals, and grown overnight at 37℃ with shaking at 240 rpm.

A set of 24-well deep well plates were then prepared with 2.5ml of Overnight

Express Instant TB Medium from Novagen and the appropriate antibiotic. These were then inoculated with 62.5μl of the overnight culture, and grown at 37℃ with shaking at 240 rpm for 3-5 hours, until the OD at 595nm was roughly 0.5. The temperature was then reduced to 25℃ and the cultures were grown overnight with shaking at 200 rpm. The cultures were then transferred back to 96-well deep well blocks, and the cells were then harvested by centrifugation for 15 minutes at 5000×g using a Beckman Coulter Avanti centrifuge with JS5.3 rotor, the supernatant then being decanted off.

Purification Cell pellets were held at -80℃ for a minimum of 20 minutes to aid efficient cell lysis, and were then thawed and resuspended in 210μl of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10mM imidazole, 1% w/v Tween 20 pH 8.0), supplemented with 1mg/ml lysozyme and 3 units/ml of Benzonase from Merck. After allowing 30 minutes for lysis to occur, lysate was collected by centrifugation for 30 minutes at 5000×g at 4℃ using a Beckman Coulter Avanti centrifuge with JS5.3 rotor. The supernatant was then transferred to microtitre plates and purified using the Qiagen BioRobot 8000 with Qiagen Magnetic Ni-NTA beads, following the manufacturer's standard protocol.

10μl of each purified protein was then mixed with 10μl of SDS PAGE loading buffer (100mM Tris pH6.8, 4% w/v SDS, 0.2% w/v bromophenol blue and 20% v/v glycerol), and then heated to 90℃ for 3 minutes to denature proteins. These were then loaded and analysed using the Novex NuPAGE SDS-PAGE Gel System from Invitrogen, following the manufacturer's instructions. Insoluble lysate also was resuspended in 8M urea buffered with 100 mM NaH_2PO_4 and 10 mM Tris to pH 8.0 to allow analysis by SDS PAGE similarly.

5.4 Protein Analysis by Western Blot

Proteins were transferred from an unstained polyacrylamide gel to an Immobilon membrane from Millipore using a BioRad Trans-Blot SD Semi-Dry Transfer Cell, and following the manufacturer's standard procedure. The standard transfer buffer was used (5.82g Trizma Base, 2.93g glycine, 200 ml methanol, made up to 1L with deionised water), polyacrylamide gels being equilibrated in transfer buffer for five minutes before being sandwiched between two pieces of blot paper soaked in the transfer buffer and run in the transfer unit at 400mA for 1 hour. BioRad Precision Plus 'Kaleidoscope' Protein Prestained Standard was used as a molecular weight marker, and to give visual verification of transfer to the Immobilon membrane. The membrane was then blocked by soaking in 3% w/v bovine serum albumin for 30 minutes to prevent non-specific binding of the antibody to the membrane.

All proteins detected using Western Blot were expressed using the GE Healthcare pGEX system, and so goat polyclonal anti-GST antibodies from GE Healthcare were used. The secondary antibody is then an anti-goat IgG horseradish peroxidase, again from GE Healthcare, allowing visualisation by a combination of chemiluminescent probe (luminol), and photographic film.

5.5 SPT Functionality Assays

Chemically competent BL21-AI were transformed with pET24a empty plasmid as a negative control, pET24a including the gene for the expression of each construct for testing, and pET16b containing the gene for the SPT from *S. paucimobilis* as a positive control, and overnight cultures grown from single colonies using the procedure described above. 150μl of each of these was then used to inoculate 15ml cultures of LB in 25ml conical flasks with the correct antibiotic. These

were then grown with shaking at 120 rpm at 37℃until their OD at 600nm was at 0.4, at which point expression was induced by the addition of IPTG to a final concentration of 1mM and L-arabinose to a final concentration of 0.2% w/v. These were then grown for a further hour with shaking at 120 rpm at 37℃ and then harvested by centrifugation at 2000×g for 10 minutes.

The cells were then resuspended in phosphate buffered saline supplemented with 2.5μCi of ^{14}C-serine (a small sample of the pellet was kept to check expression levels by SDS PAGE), and grown for a further hour with shaking at 120 rpm at 37℃. The cells were then harvested by centrifugation at 6000×g for 10 minutes, resuspended in 1ml of 10:10:3 chloroform:methanol:water, and vortexed with a small amount of glass beads for 5 minutes. The lysate was then centrifuged at 14000×g for 10 minutes, and the lower organic phase carefully extracted using a fine tipped Pasteur pipette.

Samples were then concentrated for 5 minutes in an Eppendorf 5301 VacuFuge at 30℃, and analysed using thin layer chromatography, using Merck HPTLC Silica Gel 60 plates, with chloroform:methanol:ammonium hydroxide, 40:10:1, as the solvent. The plates were then visualised using autoradiography on photographic film.

5.6 Solubility Screening

Solubility screening was undertaken as described by Jancarik in 2004 [126]. Table 6 shows the buffer conditions used for initial screening of protein stability. When more stable conditions had been identified, the use of additives such as DTT, NaCl, and glycerol was tested, and pH varied, in order to find the most stable buffer conditions for each protein.

No.	Buffer	pH
1	Glycine	3
2	Citric acid	3.2
3	PIPPS	3.7
4	Citric Acid	4
5	Sodium acetate	4.5
6	Sodium/potassium phosphate	5
7	Sodium citrate	5.5
8	Sodium/potassium phosphate	6
9	Bis-tris	6
10	MES	6.2
11	ADA	6.5
12	Bis-tris propane	6.5
13	Cacodylate	6.5
14	Ammonium acetate	7
15	MOPS	7
16	Sodium/potassium phosphate	7
17	HEPES	7.5
18	Tris	7.5
19	EPPS	8
20	Imidazole	8
21	Tris	8.5
22	CHES	9
23	CHES	9.5
24	CAPS	10

Table 6: The composition of buffers used in the optimum solubility screen. Each had a concentration of 100mM.

5.7 Crystallisation

High throughput screening was undertaken using a Screenmaker 96+8 *Xtal* from Innovadyne Technologies Inc. to produce sitting drops. This tests 96 different crystallisation solutions at a time, each with 2 drops of different concentration (Sitting drops, 100nl of crystallisation solution with 100nl of protein solution and 200nl of crystallisation solution with 100nl of protein solution). Trays were then incubated at 18℃ and any crystal growth was monitored.

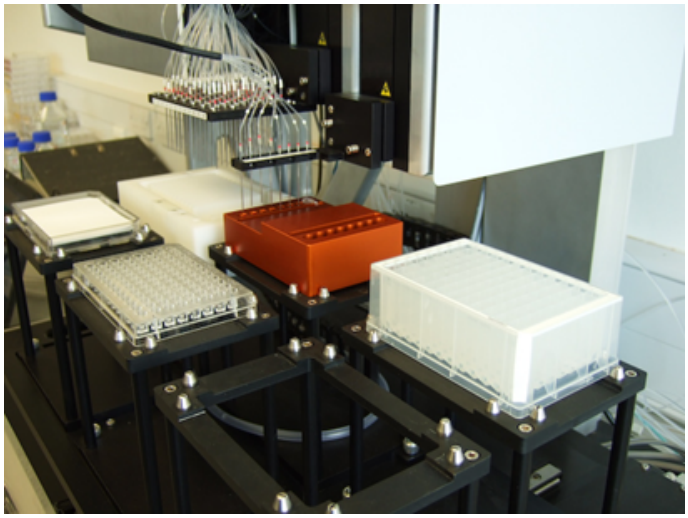


Figure 15: The Screenmaker 96+8 *Xtal* from Innovadyne Technologies Inc. is use to produce nanolitre-scale crystallisation experiments, for high-throughput screening of crystallisation conditions.

Larger scale hanging drop vapour-diffusion crystallisation experiments were set up by hand on a μl scale, with 1ml reservoirs. Drops were set up with a variety of ratios of protein solution to reservoir solution, in order to vary protein and precipitant concentration. Trays were mainly incubated at 18℃, though some were also incubated at higher and lower temperatures. Drops were monitored daily for two weeks, and less frequently thereafter. The typical crystallisation experiment used for the HSPs is shown in table 7.

0.02M	Calcium Chloride	0.02M	Calcium Chloride	0.02M	Calcium Chloride
25.0%	2-Methyl-2,4-pentanediol	25.0%	2-Methyl-2,4-pentanediol	25.0%	2-Methyl-2,4-pentanediol
0.20M	Sodium Acetate pH 5.5	0.25M	Sodium Acetate pH 5.5	0.30M	Sodium Acetate pH 5.5
0.02M	Calcium Chloride	0.02M	Calcium Chloride	0.02M	Calcium Chloride
27.5%	2-Methyl-2,4-pentanediol	27.5%	2-Methyl-2,4-pentanediol	27.5%	2-Methyl-2,4-pentanediol
0.20M	Sodium Acetate pH 5.5	0.25M	Sodium Acetate pH 5.5	0.30M	Sodium Acetate pH 5.5
0.02M	Calcium Chloride	0.02M	Calcium Chloride	0.02M	Calcium Chloride
30.0%	2-Methyl-2,4-pentanediol	30.0%	2-Methyl-2,4-pentanediol	30.0%	2-Methyl-2,4-pentanediol
0.20M	Sodium Acetate pH 5.5	0.25M	Sodium Acetate pH 5.5	0.30M	Sodium Acetate pH 5.5

Table 7: The well composition of the typical hanging drop vapour diffusion crystallisation experiment used for to the crystallisation of the heat shock proteins. It is based on the conditions published by Kim *et al.* [81]. Multiple drops were then suspended above each well, typically three, having compositions (well solution : protein solution) of 2:1, 1:1 and 1:2.

5.8 Data Collection at Durham

Initial testing of crystals was performed using Cu Kα radiation (λ=1.54184Å) on a Bruker Microstar rotating anode with a three circle goniometer and a Platinum 135 CCD area detector. ϕ and ω scans were used to sample reciprocal space for initial matrix determination. An Oxford Cryosystems Cobra was used to cool all crystals to 100K. Crystals were mounted in Hampton cryoloops with magnetic bases.

5.9 Data Collection at DLS

Synchrotron radiation datasets were collected on the macromolecular crystallography beamlines I02 and I04, and the microfocus beamline I24 at Diamond Light Source in Oxfordshire. The beamlines are based on U21 in vacuum undulators as insertion devices, crystals were mounted on a 1 axis (ϕ) goniometer perpendicular to the X-ray beam using Hampton cryoloops with a Rigaku ACTOR robot and the Uni-puck system, allowing for quick screening of samples. The sample area at I02 is shown as figure 16, and the specific details of each beamline are shown in table 8.

Name	Wavelength (Å)	Beam Size (FWHM, μm)	Detector
I02	0.68–2.48	120×80	ADSC Q315r CCD
I04	0.73–1.91	120×80	ADSC Q315r CCD
I24	0.50–1.91	5–100	Pilatus 6M

Table 8: Beamline Details: I02 and I04 are standard macromolecular beamlines, whilst I24 is a microfocus macromolecular beamline. Each beamline is based on a U21 in-vacuum undulator as the insertion device. They are equipped with robotic sample changers, and the sample stage is nitrogen-cooled to 100K as standard.

Crystals were tested by taking two frames at 90° from each other, to asses the quality of diffraction. Those showing good levels of diffraction then had full datasets collected, which were generally 180° ϕ scans with fine (generally 0.25°)

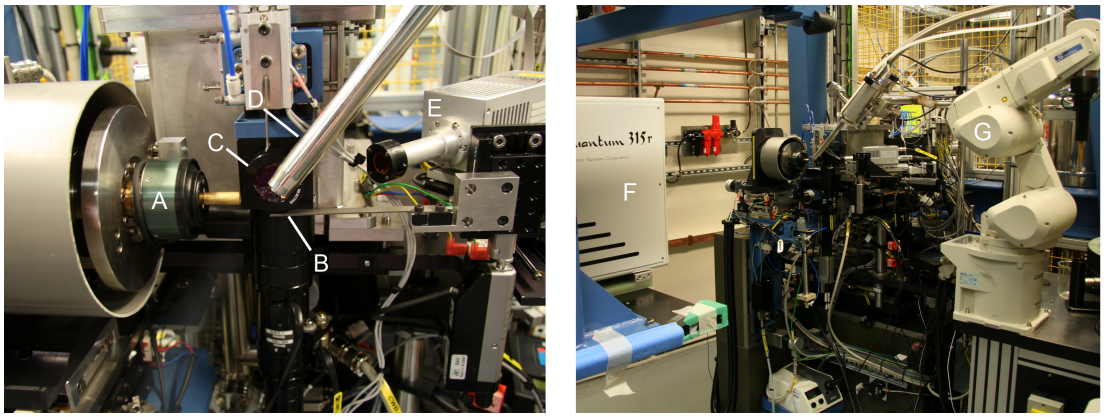


Figure 16: Macromolecular crystallography beamline I02 at Diamond Light Source: (**A**) 1-axis goniometer (ϕ); (**B**) Beamstop (retracted from experimental position); (**C**) On-axis camera for crystal centering; (**D**) Cryostream; (**E**) Fluorescence scanner; (**F**) ADSC Q315 CCD image detector (retracted from experimental position); (**G**) Robotic sample changer.

slicing. Where significant ice build up was seen on crystals, it was generally found that returning them to the sample dewar, and remounting them on the goniometer, using the robot was sufficient to remedy this.

5.10 Data Processing

In house datasets were indexed and integrated using the PROTEUM software package from Bruker. Synchrotron datasets were first indexed using the HKL2000 package before integration and scaling using the XDS package.

Structure solution and refinement were undertaken using the CCP4 package of programs [127]. Phasing was undertaken using the Phaser program and least-squares refinement used the Refmac program [128] with Coot [129] used to visualise and correct structures between successive rounds of refinements.

6 Results – Serine Palmitoyltransferase from *Toxoplasma gondii*

6.1 Bioinformatics

Bioinformatic studies of TgSTP have identified the protein to be made up of three domains. Transmembrane helix prediction using HMMTOP identified the first 47 amino acids to be a transmembrane domain. BLAST searches suggest that amino acids 181 to 571 represent a conserved PLP-dependent aminotransferase domain, showing significant homology across a broad range of species, in particular other apicomplexans. Little is known about the intervening 134 amino acids, though secondary structure prediction suggests that the domain is composed almost entirely of α-helices. A diagram showing these domains can be seen as figure 17, and a sequence alignment of the full length protein against a number of species performed using ClustalW is included as appendix B.

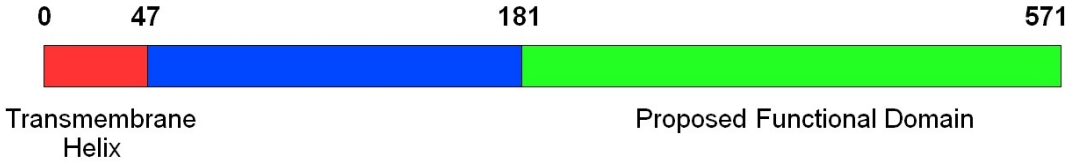


Figure 17: A diagram of the proposed domain structure of TgSPT. The transmembrane helix is shown in red, the conserved aminotransferase domain is shown in green, and the intervening 134 amino acids are shown in blue.

Whilst previous efforts had focused on the use of solubility tags to enhance expression of constructs beginning with amino acid 47, it was decided to move the study forwards by focussing on shorter constructs, aiming to find a minimal functional domain. Initial efforts concentrated on the domain with the highest homology to other aminotransferases, starting at amino acid 181. Whilst this

domain proved to give abundant expression, *in vivo* functionality studies showed that it was not functional. Therefore a number of further constructs were designed based on areas predicted to be random coil between α-helices, starting at amino acids 55, 69, 77, 90 107, 120, 132, 144, 159 and 177. Along with 47 and 181, these were used for high-throughput screening at the Oxford Protein Production Facility.

6.1.1 Comparative Modelling

The homology model of the conserved functional domain of TgSPT was calculated using MODELLER [124]. The complete SPT sequence was used in the most exhaustive search for similar sequences and resulted in a top hit with 29% sequence identity with the crystal structure of SPT from *Sphingobacterium multivorum* (PDB code: 3A2B) determined with the substrate L-serine covalently bound to the co-factor PLP [130]. Other hits included the closely related SPT from other sphingobacteria which showed the same overall fold albeit with lower sequence identities. MODELLER then automatically calculated a model containing all non-hydrogen atoms, and implemented comparative protein structure modeling by satisfaction of spatial restraints and modelled loops in the protein structure *de novo* [131, 132]. All models were then analysed and least-square superpositions were performed with Coot [125]. The result is shown in figure 18, and shows that, as might be expected, the protein is likely to be of a similar fold to the other known SPT structures.

6.2 pETzt2_1

Previous work had suggested difficulties in extracting soluble protein [37]. For this reason the vector pETzt2_1 was chosen (Figure 10). It includes a 10kDa solubility enhancing and multi-histidine fusion tag linked to the protein by the

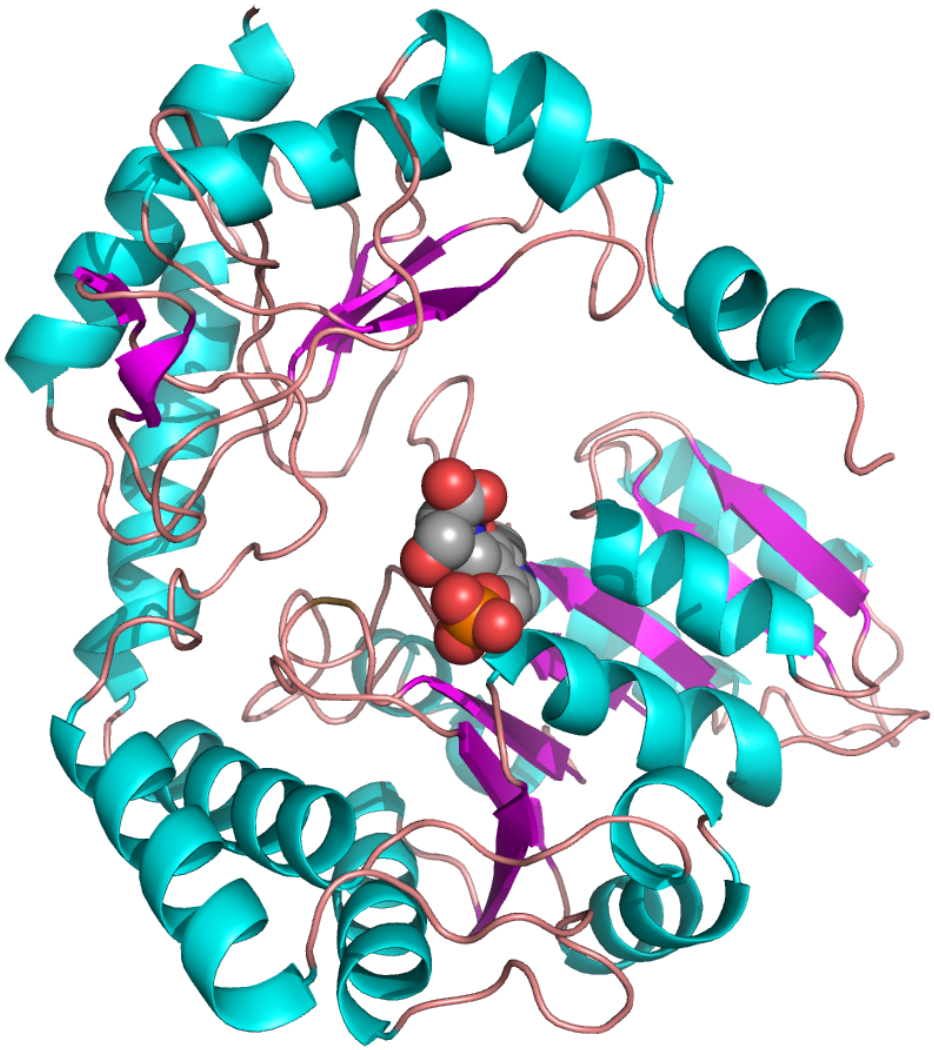


Figure 18: The figure shows the overall fold of one SPT monomer with the PLP-serine position taken from from 3A2B which clearly identifies the open binding cleft. α-helix is shown in cyan; β-sheet in magenta; and random coil in puce. It is important to note that although the overall position of the co-factor is very likely to be similar the active site residue (Lys265) is not conserved in SPT from *T. gondii* (see sequence alignment) . In solution the enzyme is likely to be a closely inter linker homo-dimer as seen in other family members [130, 36].

recognition site for the TEV protease allowing for later cleavage of the tag to give the desired construct for crystallisation. The sequences of the primers used are shown in Table 2. The sequence of the full construct, including the tag, is included in appendix A. Masses, theoretical pI values for the tagged and cleaved protein were calculated using the online ProtParam and are shown in Table 9

	Tagged Protein	**Cleaved Protein**
Number of Amino Acids	484	395
Molecular Weight\gmol^{-1}	53777.3	43623.2
Theoretical pI	6.84	8.11

Table 9: Parameters calculated for TgSPT:181 from pETzt2_1 before and after fusion tag cleavage.

6.2.1 Ovexpression

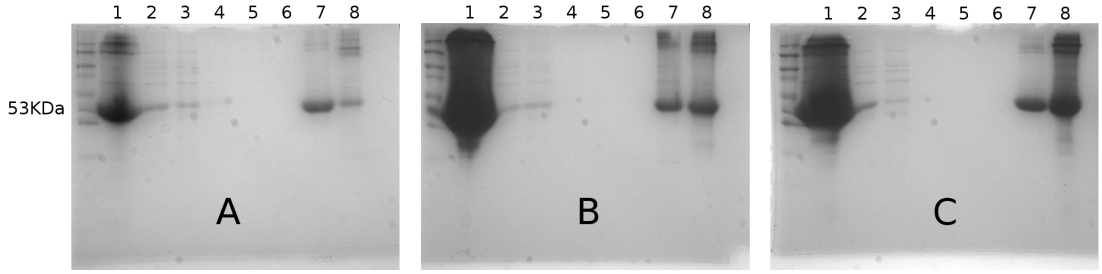


Figure 19: Expression and purification of tagged TgSPT181 from pETzt2_1 at three temperatures: **A** 20℃; **B** 30℃; **C** 37℃. Key to lanes: **1** Total cell lysate; **2** Pre-incubation soluble; **3** Post-incubation soluble; **4-6** Wash fractions; **7** Purified protein; **8** Total insoluble protein.

Initial Testing Initial expression tests were performed at 20℃, 30℃ and 37℃. Small scale purification was undertaken for each sample which were then analysed by sodium docecylsulphate polyacrylamide gel electrophoresis (SDS PAGE) (Figure 19). Good levels of expression were found at each temperature with the highest levels of soluble protein at 30℃.

6.2.2 Purification

Quantities (∼15mg) of tagged TgSPT have been isolated and purified as shown in figure 20. Tag cleavage was attempted using TEV protease. As shown by the lack of cleaved protein in lane **5** of the gel in Figure 20 this was unsuccessful, leaving only tagged protein after overnight incubation at 4℃. It is believed that this is due to steric inhibition, due to an insufficiently long linking region between the cleavage site and the protein of interest. A positive control to test the activity of the enzyme was performed by the expression of green fluorescent protein using an identical system, from which the tag was successfully cleaved. This lends further weight to the suggestion that it is a steric effect. However, whilst it is rather difficult to imagine an alternative cause of the unsuccessful tag cleavage, it would be rather difficult to prove this out-right, as this would require the re-cloning of the construct with a longer linking region. Further difficulties were also had with medium-term protein stability in solution, though these were resolved by changing the buffer conditions for the protein.

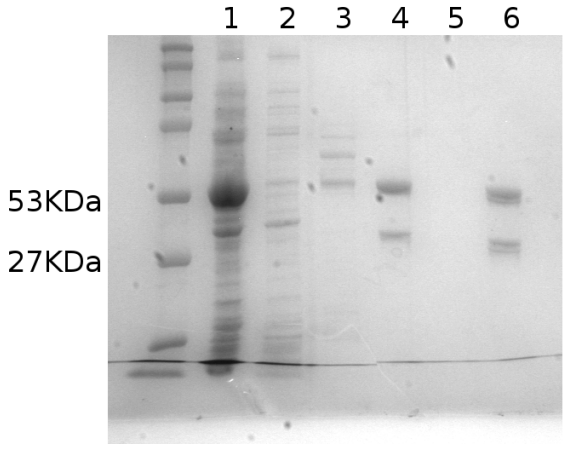


Figure 20: Large scale expression and purification of tagged TgSPT181 from pETzt2_1. Key to lanes: **1** Total cell lysate; **2** Initial pre-incubation soluble; **3** Initial post-incubation soluble; **4** Purified protein with TEV protease after overnight incubation; **5** Unbound protein; **6** 500mM imidazole column wash.

6.2.3 Protein Buffers

A number of buffer conditions were screened across various values of pH with a variety of different additives and buffers. An efficient optimum solubility screen as outlined by Jancarik *et al.* [126] was carried out, though with only qualitative evaluation of the levels of precipitation, as low protein concentration and heavy precipitation made the use of dynamic light scattering unfeasible. This qualitative evaluation was made by placing the drops of protein solution under a microscope, and making a comparative judgement of in which condition there was the least precipitated protein. pH was then further screened in stable conditions, and the use of additives tested. Protein was found to be most stable in a 100mM sodium/potassium phosphate buffer at pH 7.4 (made by adding 100mM sodium phosphate to 100mM potassium phosphate until pH 7.5 is reached) with 1mM dithiothreitol (DTT), 10% glycerol and 200mM NaCl, with DTT being the key ingredient for maintaining solubility, suggesting that earlier problems had been due to oxidation of the protein in solution.

6.2.4 PLP Binding

PLP has a strong yellow colour, and absorbs light strongly in the UV-blue region of the electromagnetic spectrum. The OD at 400nm of purified TgSPT181 from pETzt2_ 1a was measured before and after dialysis in buffer containing PLP, and showed significant increases in absorption, suggesting that the PLP binding site was intact in this heavily truncated construct.

6.3 pET24a

Previous work [37] had produced untagged constructs of full length TgSPT and TgSPT47 in pET24a, which were shown to be functional *in vivo* by radio-isotope

labelling, using the same protocols as have been used for the work outlined here.

An analogous construct for TgSPT181 has now been produced, and shown to give high levels of soluble expression. The previous functionality study was then repeated including this construct. The autoradiograph is shown as figure 21, and clearly shows that no radiolabelled 3-dehydrosphinganine was produced by this construct TgSPT181. Following this result, it was decided to return to tagged longer constructs, with the aim of producing protein which was both soluble, and functional.

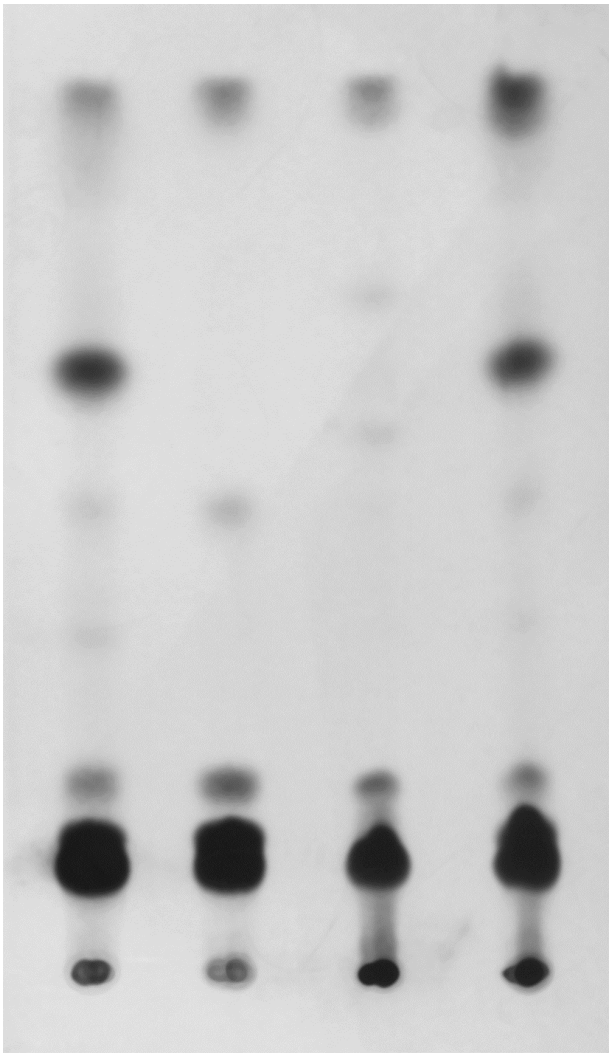


Figure 21: An autoradiograph of a TLC plate of lipids extracted from *E. coli* cells containing two truncations of TgSPT as well as SpSPT as a positive control, which were grown in radiolabelled serine. The dark spot clearly visible 2/3 of the way up the plate on the first and fourth lanes is 3-dehydrosphinganine. Lanes (from left to right) 1: TgSPT47; 2: TgSPT181; 3: Empty vector (pET24a); 4: SpSPT.

6.4 pGEX6.2

GST-tagged TgSPT47 was prepared using the vector pGEX6.2, and expressed from a variety of cell lines expressed at both 16 and 30℃, and with two different concentrations of IPTG (0.1 and 1mM). Crude soluble lysate from small-scale expression trials did not show obvious expression bands with SDS PAGE, so an anti-GST Western Blot was performed (figure 22), which clearly shows the presence of soluble GST-TgSPT47 fusion at around 80KDa from a number of conditions.

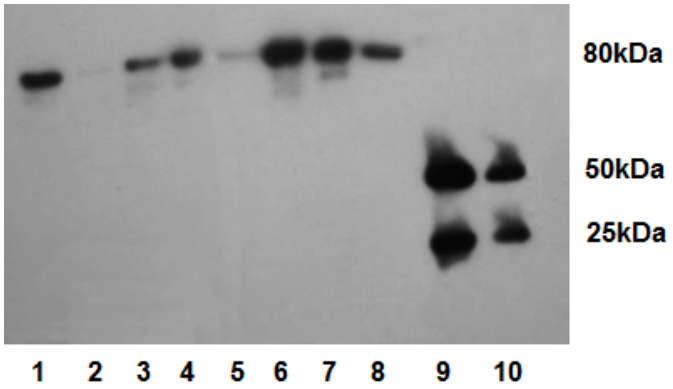


Figure 22: Western blot showing soluble expression of GST-TgSPT47 from pGEX6.2 at c. 80kDa. Lanes 1–4 were expressed from BL21(DE3) and lanes 5–8 from Tuner. Lanes 9 and 10 contain GST as a positive control, the two bands corresponding to roughly 25 and 50kDa; the monomer and homodimer.

Whilst little correlation was visible in general between the conditions, it was clear that Tuner cells with higher concentrations of IPTG at 30℃ gave the most consistently strong bands. Upon scaling up these conditions, the quantity of soluble protein was seen to reduce greatly, perhaps due to a lengthened period of expression causing overproduction of the protein, and subsequent production of inclusion bodies in the cells. As with other constructs, whilst there seems to be a theoretical possibility of producing soluble TgSPT47, in practice larger scale cultures do not yield the expected result, and only vanishingly small quantities of protein can be isolated.

6.5 High-throughput Protocols

At least one successful construct was cloned for all but three of the 48 constructs designed for high-throughput screening. This is shown by figure 23, which is a agarose gel of the verification PCR carried out on the 96 plasmids prepared (two for each construct). The gel was run from left to right in the image shown. As constructs are all of a similar length, it is easy to see those which did not work, as they are significantly shorter than those around them (such as lanes D1, E7 and E12) suggesting that a construct shorter than the length intended was cloned, or no strong band is visible at all (such as lane C1) suggesting that cloning was completely unsuccessful.

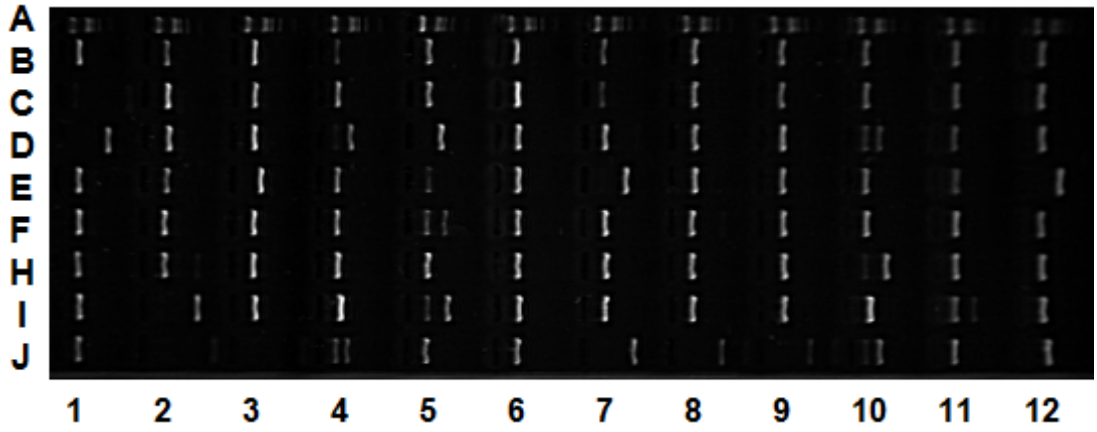


Figure 23: An OPPF gel showing verification PCR. Lanes A1–12 and contain DNA length standards. The verification PCR for first version of each construct runs from J2 through to B6, in the order shown in tables 10 & 11, and the second version from J7 through to B12.

These were then expressed by IPTG and autoinduction from B834 and Rosetta(DE3)pLysS. Two gels showing protein purified from pOPINJ and pOPINS3C grown in Rosetta(DE3)pLysS, by both autoinduction and IPTG induction are shown as figure 24. The proteins of interest would be at around the level of the 75kDa molecular weight marker. On the top gel (from autoinduction), whilst there is some limited expression seen for a number of conditions, significant expression is only seen in the last four lanes.

On the bottom gel (from IPTG induction), more expression is seen in lanes 10, 16, and 18-25. The drop in the vertical position of the band from lane 16 to 25 is due to the steady decrease in the length of the cloned construct. The right hand side of bottom gel shows the most successful set of conditions: pOPINS3C from Rosetta(DE3)pLysS by IPTG induction. A full list of the conditions trialled, and their success are shown in tables 10 & 11.

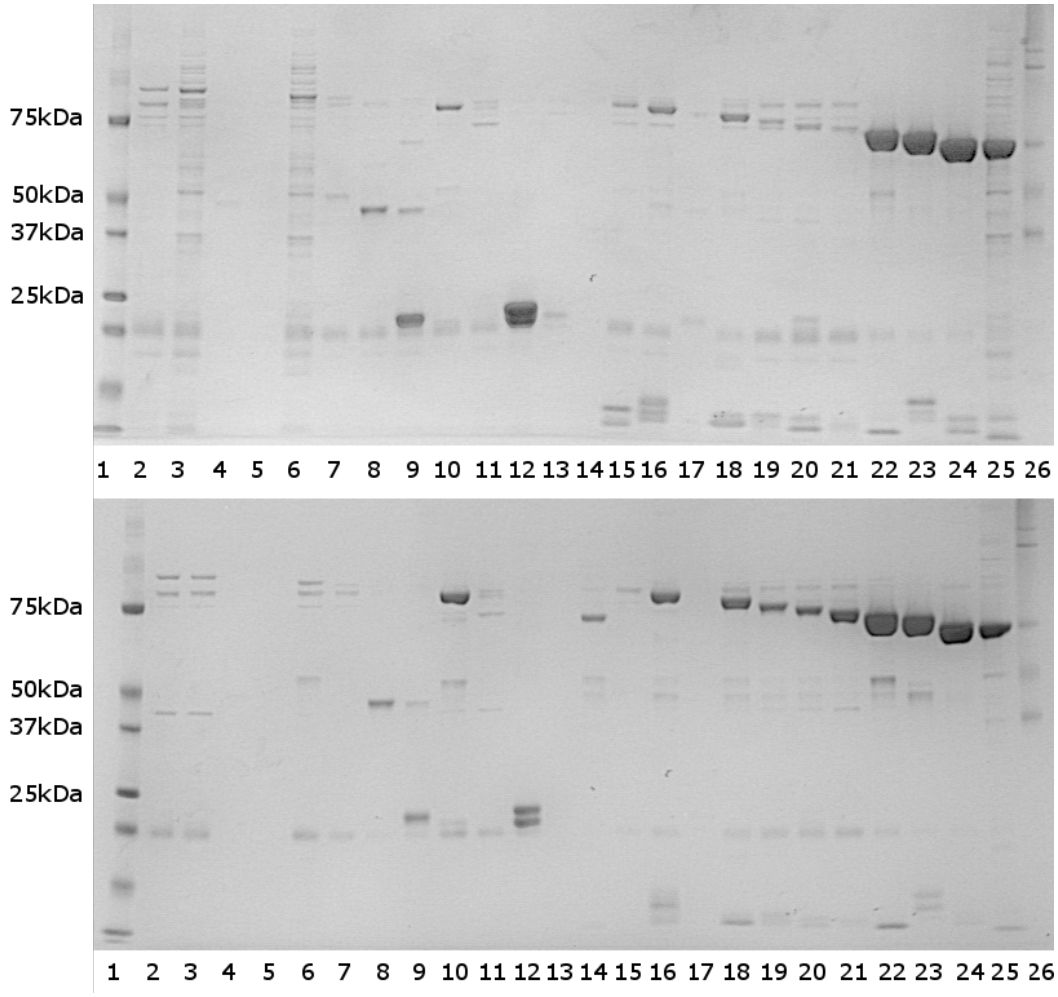


Figure 24: OPPF Gels: Expression from pOPINJ (lanes 2–13) and pOPINS3C (lanes 14–25) in Rosetta(DE3)pLysS: autoinduction (top); 0.5mM IPTG induction (bottom). Lane 1 contains Precision Plus Protein Standard from BioRad.

Perhaps the most surprising result, given previous success with expression from pGEX6.2 of GST tagged TgSPT47, is the distinct lack of success in the expression

of GST tagged protein from pOPINJ. Another interesting result is that the N-terminal His_6 in pOPINF performed far better than the C-terminal His_6 tag in pOPINE, which are otherwise identical. This might suggest that the C-terminal of the protein is not readily accessible from the surface, and that C-terminal tagging has an effect on the correct folding, and therefore solubility, of the protein.

Vector	First AA	Final AA	Condition 1	Condition 2	Condition 3	Condition 4
pOPINF	48	571				
pOPINF	55	571				
pOPINF	69	571				
pOPINF	77	571				(+)
pOPINF	90	571	(+)		+	(+)
pOPINF	107	571				(+)
pOPINF	120	571	(+)	(+)		+
pOPINF	123	571	+	++	(+)	(+)
pOPINF	144	571	++		(+)	
pOPINF	159	571	+	++	+	+++
pOPINF	177	571	+++	+++	+++	+++
pOPINF	181	571	+++	++	+++	+
pOPINE	48	571				
pOPINE	55	571				
pOPINE	69	571				
pOPINE	77	571	+	(+)		(+)
pOPINE	90	571	+	(+)	+	
pOPINE	107	571				
pOPINE	120	571			+	
pOPINE	123	571			+	
pOPINE	144	571	(+)	++		
pOPINE	159	571	+			++
pOPINE	177	571	++	++	++	+
pOPINE	181	571	++	++	+++	++

Table 10: Success of preliminary expression and purification testing of constructs expressed at OPPF under various conditions. Key to conditions: **1**: Autoinduction, B834(DE3); **2**: Autoinduction, Rosetta(DE3)LysS; **3**: IPTG, B834(DE3); **4**: IPTG, Rosetta(DE3)LysS. Key to symbols: (+): Some limited expression; +: Reasonable expression; ++: Strong expression; and +++: Expression saturating Ni-NTA beads.

Vector	First AA	Final AA	Condition 1	Condition 2	Condition 3	Condition 4
pOPINJ	48	571				
pOPINJ	55	571				
pOPINJ	69	571				
pOPINJ	77	571				
pOPINJ	90	571				
pOPINJ	107	571				
pOPINJ	120	571			(+)	
pOPINJ	123	571			(+)	
pOPINJ	144	571	+	++		(+)
pOPINJ	159	571	+			
pOPINJ	177	571				
pOPINJ	181	571				
pOPINS3C	48	571	(+)			
pOPINS3C	55	571	(+)			(+)
pOPINS3C	69	571		++		+
pOPINS3C	77	571				
pOPINS3C	90	571	+	+		++
pOPINS3C	107	571		(+)		+
pOPINS3C	120	571		(+)		+
pOPINS3C	123	571	(+)	(+)	+	++
pOPINS3C	144	571	+	+++	+++	+++
pOPINS3C	159	571	+	+++	++	+++
pOPINS3C	177	571	++	+++	+++	+++
pOPINS3C	181	571	++	++	++	++

Table 11: Success of preliminary expression and purification testing of constructs expressed at OPPF under various conditions. Key to conditions: **1**: Autoinduction, B834(DE3); **2**: Autoinduction, Rosetta(DE3)LysS; **3**: IPTG, B834(DE3); **4**: IPTG, Rosetta(DE3)LysS. Key to symbols: (+): Some limited expression; +: Reasonable expression; ++: Strong expression; and +++: Expression saturating Ni-NTA beads.

6.6 Summary

Whilst bioinformatics have suggested that a minimal functional domain exists beginning at around amino acid 181, *in vivo* functionality testing using protein expressed from *E. coli* have shown that this minimal construct does not show functionality. A variety of constructs were tested, using both low-throughput and high-throughput protocols. The best results were found to use short, sharp expression with induction by IPTG from pET28a 3C-protease-cleavable His_6 and a SUMO tag.

7 Results – Small Heat Shock Proteins from *Methanococcus jannaschii*

7.1 Bioinformatics

Bioinformatics and construct design were undertaken by Andrew Landsbury in Durham's Department of Biological and Biomedical Sciences. An alignment of MjHSP16.5 with human αB-crystallin performed with ClustalW is included as appendix G. Two C-terminal truncation mutants, 3FR and 4FR, designed to mimic the myopathy-causing Q151X truncation in αB-crystallin., were produced. The extent of the truncation in 4FR is shown in figure 25, and their sequences included as appendix F. A third mutant, R107G, was designed to mimic the myopathy causing R120G mutation in αB-crystallin.

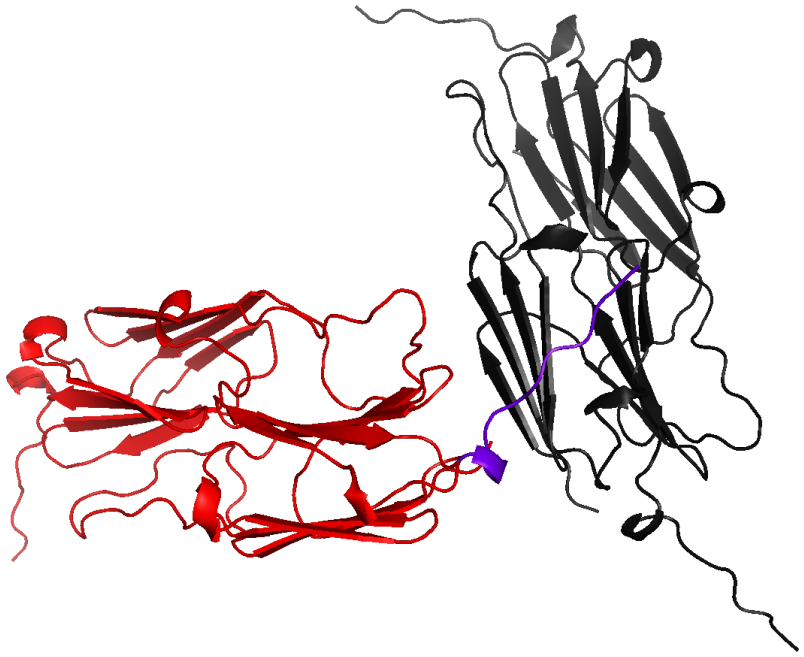


Figure 25: The interface between two dimers of HSP. The purple section shows the extent of the 4FR truncation, on the red dimer.

7.2 Protein Production

Native and truncated proteins were expressed and purified by Andrew Landsbury in the Department of Biological and Biomedical Sciences, Durham University following the procedures laid out by Kim *et al.* [81]. SDS PAGE was then used to asses the purity of samples (an example is shown as figure 26), and LC FTICR MS was then used to verify the identity of the protein by its exact mass (an example used to identify that R107G had been correctly mutated is shown as figure 27).

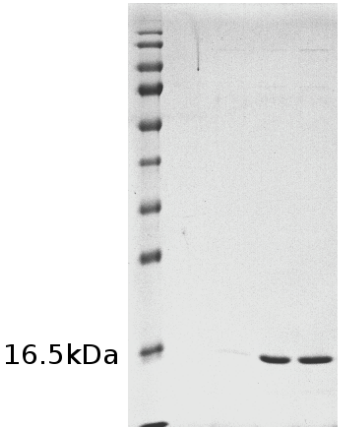


Figure 26: An SDS PAGE gel showing the purity of samples of native MjHSP used for crystallisation.

7.3 Native Protein

Positive results were observed in two drops and selected for further screening: Conditions 34 (35% (v/v) MPD, imidazole pH8.0, 0.2M $MgCl_2$) and 37 (40% (v/v) PEG-300, MES pH6.5) of the Emerald BioSystems Cryo I+II screen. A good yield of crystals was found in the condition 37 optimisation, though it was not possible to reproduce the crystal growth from condition 34.

Further trays were also set up using the conditions outlined by Kim *et al.* and incubated at both 4℃ and 18℃, as shown in table 7. These yielded a significant number of good sized crystals which grew overnight in all drops at 18℃, though

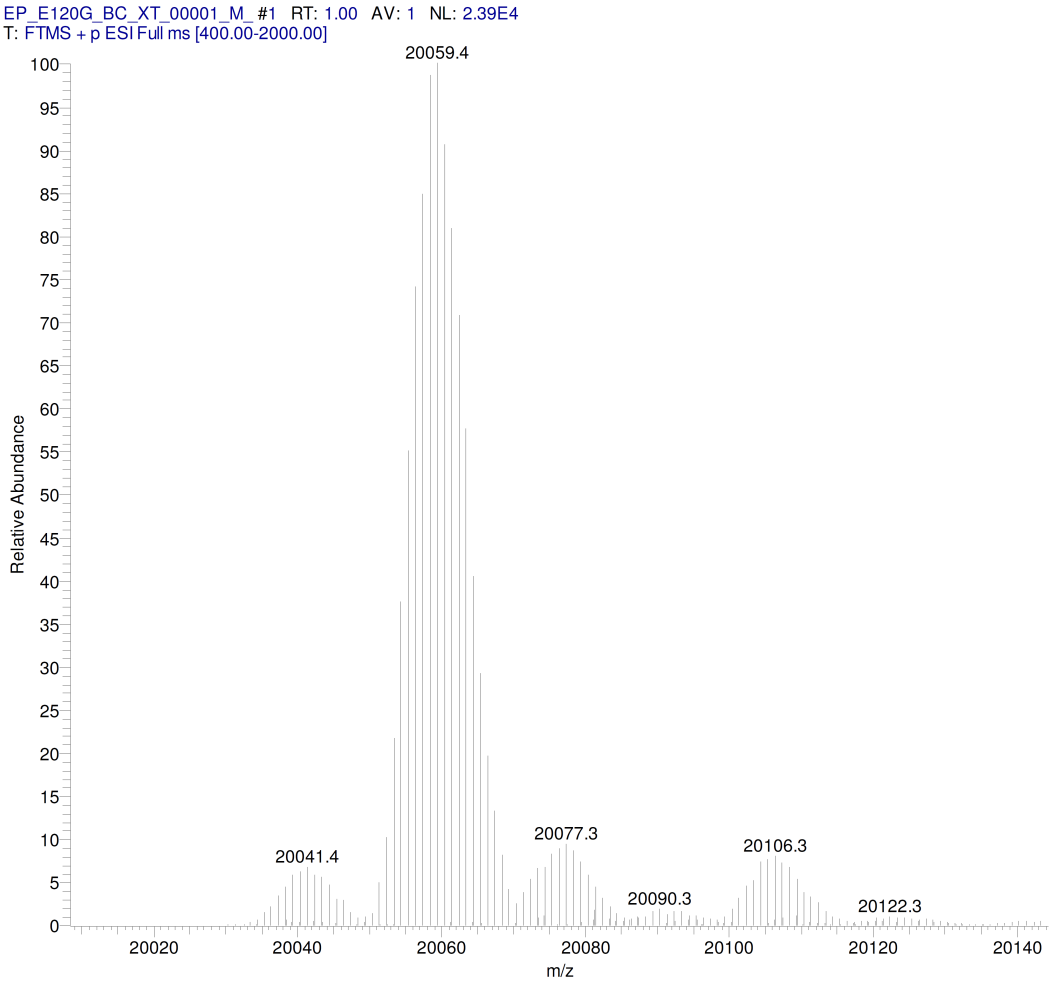


Figure 27: The deconvoluted mass spectrum used to verify that the R107G mutation had been successful. The mass peak at 20059.4gmol^{-1} corresponds well to the expected mass of 20059gmol^{-1}.

far fewer crystals were seen to grow at 4℃, and those which did required weeks to grow. Diffraction was observed to approximately 6.5Å for the crystals grown in screen condition 37. No further testing was possible due to technical difficulties with the diffractometer.

Further crystals were grown, and tested on beamline I02 at Diamond Light Source. Diffraction was seen to ∼2.45Å for the crystals grown from the published conditions and, less successfully, to 2.8Å for the crystals from the condition 37

Protein	Date	Beamline	Resolution	Frames	ϕ Slicing	ϕ Range
MjHSP	15-Apr-2009	I02	2.45Å	360	0.5°	180°
3FR	15-Jul-2009	Durham	7.40Å	623	0.1°	62.3°
3FR	27-Sep-2009	I02	3.85Å	180	0.5°	90°
3FR	27-Sep-2009	I02	3.95Å	180	0.25°	45°
3FR	23-Feb-2010	I24	3.35Å	300	0.4°	120°
3FR	23-Feb-2010	I24	3.20Å	300	0.4°	120°
4FR	09-Jun-2009	Durham	7.80Å	605	0.1°	60.5°
4FR	27-Sep-2009	I02	3.45Å	180	0.5°	90°
R107G	23-Feb-2010	I24	3.20Å	225	0.4°	90°
R107G	23-Feb-2010	I24	3.35Å	225	0.4°	90°
R107G	23-Feb-2010	I24	3.15Å	225	0.4°	90°
R107G	17-Dec-2010	I04	3.25Å	180	0.25°	45°
R107G	22-Jan-2011	I04	2.60Å	520	0.35°	182°
R107G	22-Jan-2011	I04	3.00Å	182	1°	182°

Table 12: Full datasets collected on crystals wild-type MjHSP16.5, R107G, and the truncation mutants.

optimisation. A 180° ϕ scan with a slicing of 0.5° was collected at 1 second per frame. A typical frame, showing diffraction to ∼2.5Å is shown in figure 28.

The unit cell was found to be the same as that for the published structure (Space group: $R3$, $a = b = 171.58$Å, $c = 102.22$Å). Further data collection statistics are shown in table 13. It should be noted, that whilst the extension in resolution against the published structure does not sound all too profound, it represents the collection of 50% more data; 33,049 reflections as opposed to the 22,008 recorded for the structure published by Kim *et al.*.

The data were indexed and integrated using HKL2000 and XDS. Structure solution was then undertaken using Phaser by molecular replacement with the published structure (PDB code: 1SHS). A number of different levels of NCS restraint including differing restraints for the main chain and side chain atoms were tested, and in both cases the tightest level of restraint was found to give the best results, with the lowest values of R and R_{free}. A representative section of the electron density is shown in figure 29, giving an idea of the extent to which the

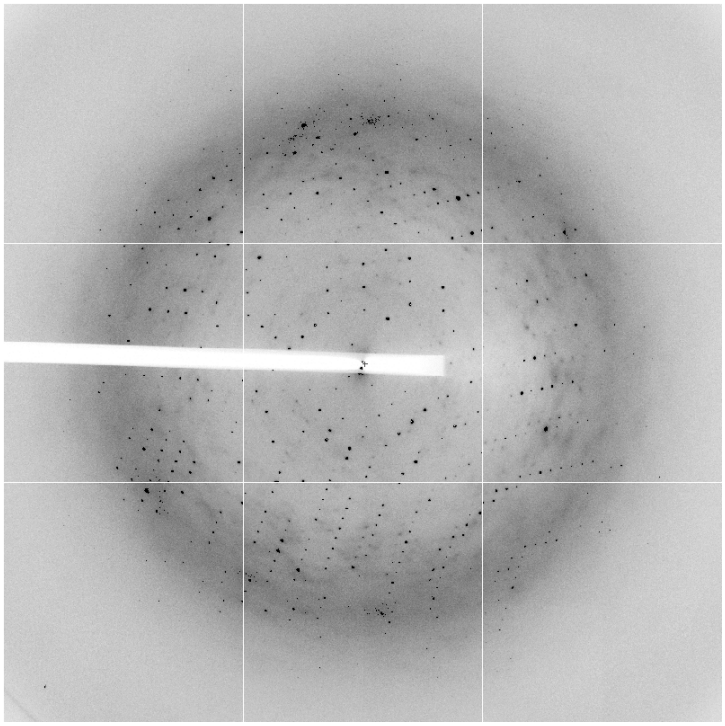


Figure 28: A typical frame of diffraction collected on crystals of MjHSP16.5 at Diamond Light Source showing diffraction to ∼2.5Å resolution.

secondary structure can be traced through the density. A number of side chains were found to be disordered and treated as having an occupancy of zero.

Table 13: Comparative data collection and refinement statistics for MjHSP.

	Published Structure	**Higher Resolution**
Data Collection Statistics		
Resolution\Å	30.0–2.9	85–2.5
Completeness	99.7%	98.6%
R_{sym}	0.037	0.034
Refinement Statistics		
Unique reflections ($F > 2\sigma$)	22,008	33,049
Completeness ($F > 2\sigma$)	80.6%	85.1%
R factor	0.216	0.252
R_{free}	0.251	0.272
Bond-length deviation\Å	0.023	0.017
Bond-angle deviation	1.505°	1.65°

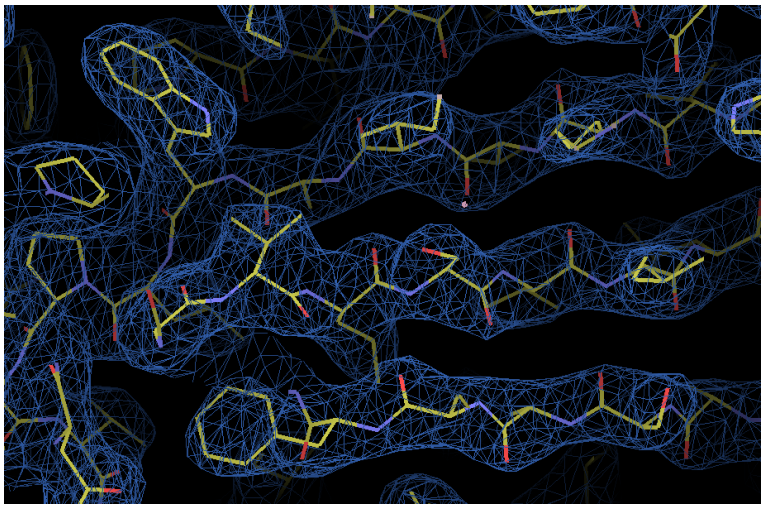


Figure 29: Representative region of the $2F_o$–F_c electron density map based on the redetermined structure contoured at 1.22σ.

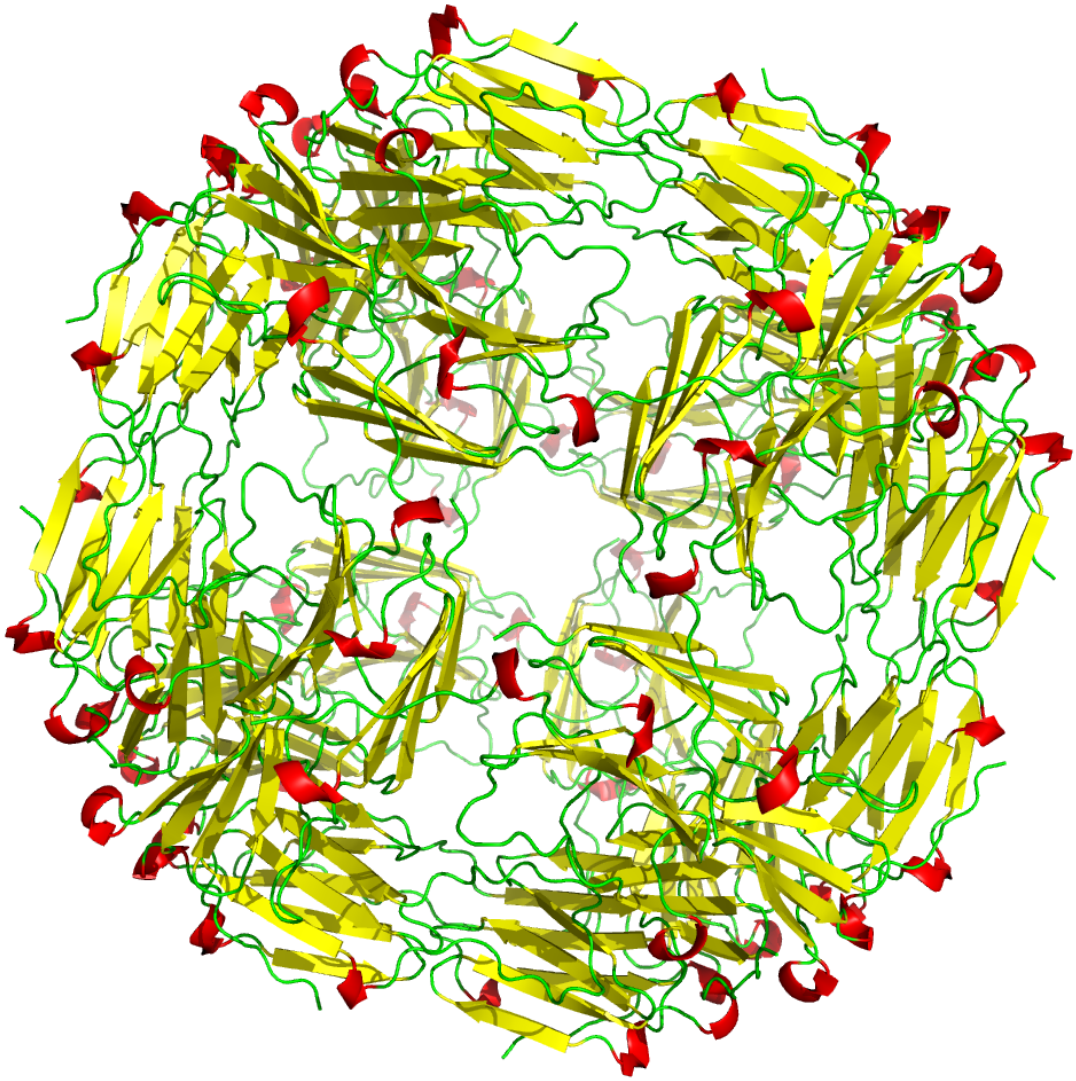


Figure 30: Cartoon depiction of the higher resolution structure of the icositetramer of MjHSP, looking down a four-fold axis. α-helix is shown in red; β-sheet in yellow; and random coil in green.

7.4 Truncated Proteins

A crystallisation screen was set up for truncate 3FR under the published conditions for the full length protein. High levels of precipitation were noted, though after a week of incubation at 18℃ small, 0.05 × 0.02 × 0.01mm, crystals were seen to form in all drops. Further screens were set up, and larger crystals suitable for diffraction studies were obtained. Initial testing in house showed only limited diffraction to 7–8Å. More crystals were grown, and taken to DLS, where diffraction was observed to 3.85Å. Initial indexing suggests the crystal to have a tetragonal unit cell of ∼125×125×180Å.

Further rounds of optimisation were undertaken, and Andrew Landsbury improved the purification protocol to give higher purity protein, though significant increases in concentration were not possible, due to heavy precipitation on spin concentration. This yielded increased size crystals, though nowhere near the size of those obtained for the wild type protein, at around $0.15 \times 0.05 \times 0.05$mm. These were taken to the microfocus beamline I24 at Diamond Light Source, where diffraction to 3.20Å was observed, though at the time of writing it has not been possible to extend this resolution further. Details of all full datasets collected are included in table 12.

High levels of non-crystallographic symmetry are shown by the auto-rotation function shown in figure 31. This function is created by taking one Patterson synthesis and rotating it by a set angle, each peak in the function representing the extent of coincidence between the rotated and original functions at the relevant degree of rotation. For example, in the function shown for Chi=120, the four peaks each represent a separate 3-fold axis of non-crystallographic symmetry. This high level of non-crystallographic symmetry has also proved a significant hindrance in structure solution using molecular replacement, as it creates multiple solutions to the equations, giving uncertainty in the result. A number of different models based

on the wild type structure were tried for molecular replacement aiming to find a model which represents the contents of the asymmetric unit, but without success. It is however clear from the high level of non-crystallographic symmetry present that the functional unit of an octahedral icositetramer must be maintained, and therefore that only crystal packing is significantly affected by the truncation.

As with truncate 3FR, a crystallisation screen was set up for truncate 3FR under the published conditions for the full length protein. Again high levels of precipitation were noted, as well as the growth of small crystals. Further to this a number of slightly larger crystals , $0.1 \times 0.05 \times 0.02$mm, grew in one drop at the highest concentration of precipitant. These were tested on the rotating anode and diffraction was observed to ∼6Å. Figure 32 shows one such crystal mounted in a cryoloop along side a frame showing the extent and quality of its diffraction. A 60° ϕ scan with a fine slicing of 0.1° was collected at 120s per frame. Initial indexing suggested the crystal may have a tetragonal unit cell of ∼125×125×180Å, as with 3FR.

These results were confirmed with larger crystals of 4FR, which were tested at Diamond Light Source. Given the strong similarities seen between the datasets for 3FR and 4FR, it was decided to discontinue study of 4FR, and focus efforts on obtaining better crystals of 3FR.

7.5 Point Mutation R107G

As with other mutants, conditions were screened around those published for the wild type protein. Improvement in crystal quality through successive rounds of improving the purity of the protein, and optimisation of crystallisation conditions can be seen in figure 33. Initially arrays of very small crystals were grown. It was then possible to improve these to small single crystals, and then to large single crystals, of up to 1mm in their longest dimension. The main factor driving

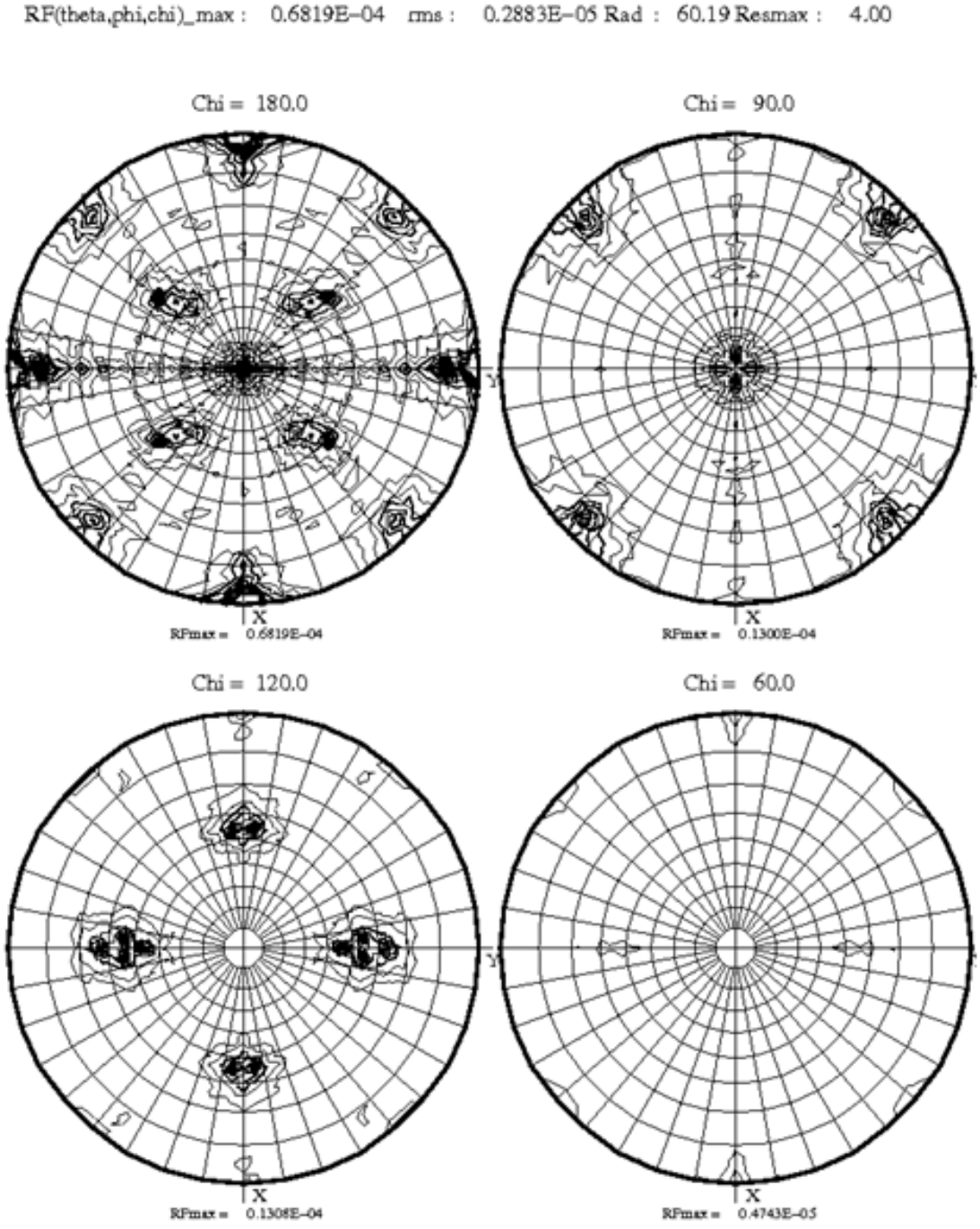


Figure 31: Auto-rotation function calculated for a 3FR dataset, showing the high level of non-crystallographic symmetry present [133]. 8 peaks can be seen corresponding to 2-fold NCS, 4 corresponding to 4-fold NCS, and 4 corresponding to 3-fold NCS.

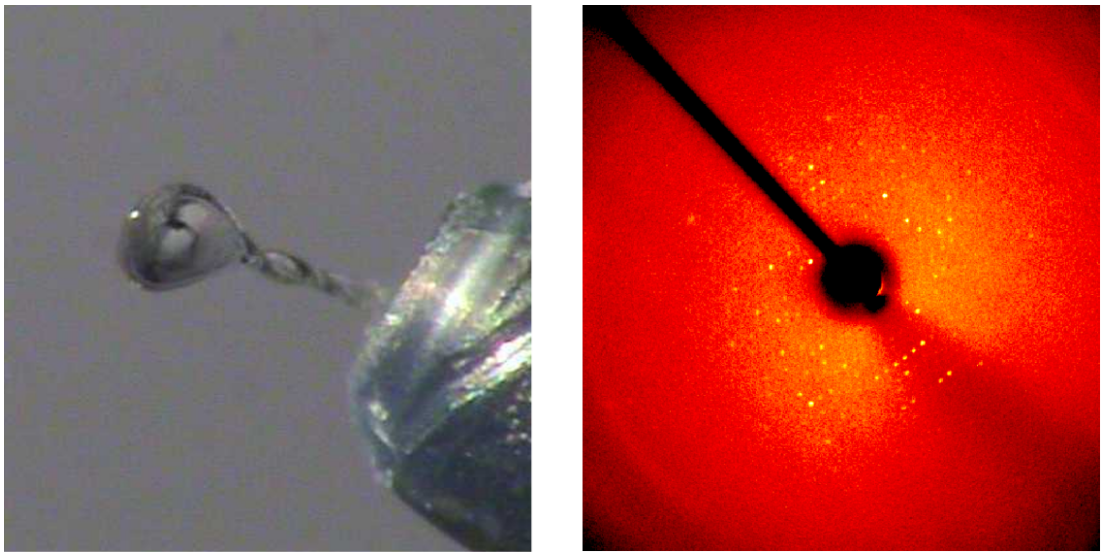


Figure 32: A crystal of mutant 4FR mounted in a 0.2mm cryoloop and a frame showing its diffraction (0.1° ϕ scan, 120s, weak diffraction visible to ~6Å).

forwards improvements in crystal quality was once again the improvement in the purity and concentration of the protein solution. Concentrations of up to 40mg/ml were used, though it was generally found that 20–30mg/ml yielded the best results.

Figure 34 shows a very large crystal of R107G mounted in a LithoLoop from Molecular Dimensions on beamline I24 at Diamond Light Source. Due to the size of the crystal, and relatively poor back-lighting, it is not possible to see the crystal fully. However, interestingly, another factor which compounds this is the general opacity of the crystal. This seems to have been due to the crystals growing around the amorphous precipitate, and yet does not seem to have had a particularly adverse effect on the general quality of diffraction. Indeed the strongest trend in the quality of diffraction was a direct correlation between the resolution obtained and the size of the crystals. It can only be assumed that the factors allowing such large crystals to grow, also increase the general intrinsic quality of diffraction.

After many rounds of optimisation, a 2.85Å dataset was collected on beamline I24 at Diamond Light Source (details of all full datasets collected are included

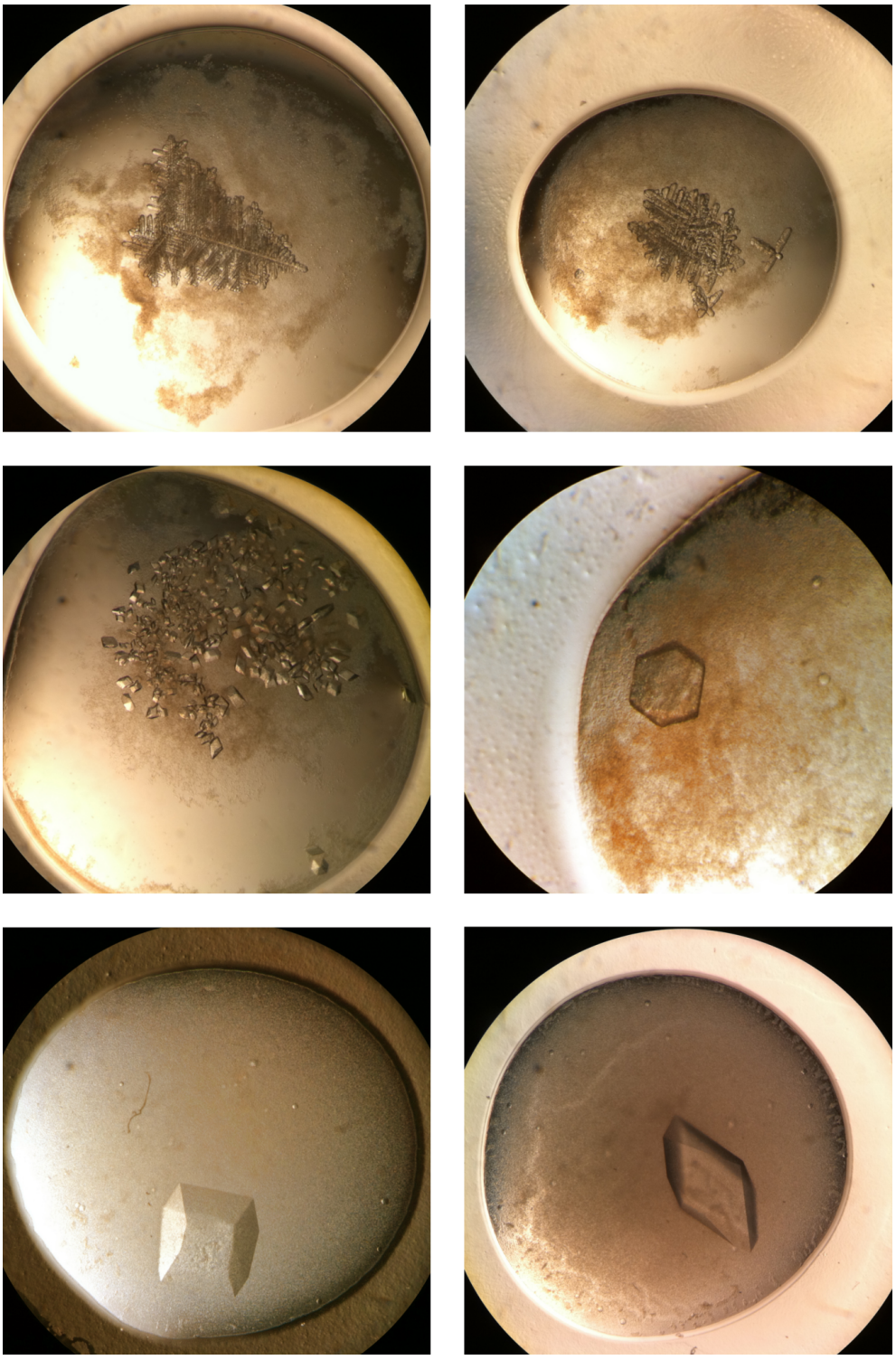


Figure 33: Various images of hanging drops containing crystals of R107G, showing progressive improvements in the quality of crystals. The largest single crystals are of the order of 1mm along their longest axis.

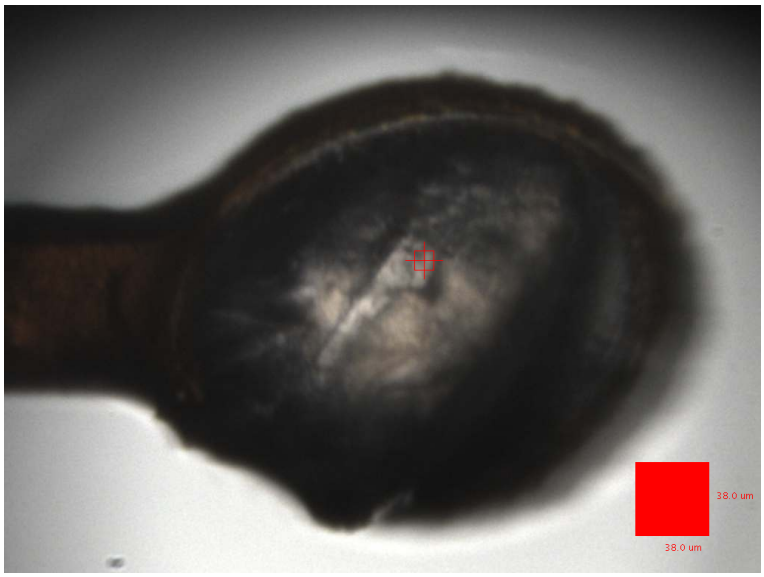


Figure 34: A large crystal of R107G mounted on the microfocus beamline I24 at DLS. The small red square shows roughly the position and the size of the beam used for data collection.

in table 12). The data were then indexed and integrated using HKL2000, and phasing was undertaken using Phaser, using the published wild type structure as a model. the unit cell was found to be near identical to that for the wild type structure (Space group: $R3$, $a = b = 173.60$Å, $c = 103.00$Å). The structure of R107G imposed upon the wild type structure is shown in figure 35, from which the broad similarities between the two structures are clear. Furthermore, the calculated route mean squared deviation (rmsd) on the C^{α} atoms between the wild type structure and R107G, for each chain are between 0.3 and 0.4Å, well within the standard experimental error, more conclusively demonstrating that there is no significant difference between the two structures.

For phasing, the arginine was included in the model, and therefore a significant region of negative electron density can be seen in the density map, as shown in figure 36. The arginine was then replaced with glycine, and the significantly cleaner electron density after refinement is shown in figure 37. The crystallographic refinement therefore quite unambiguously shows that the arginine has been mutated to

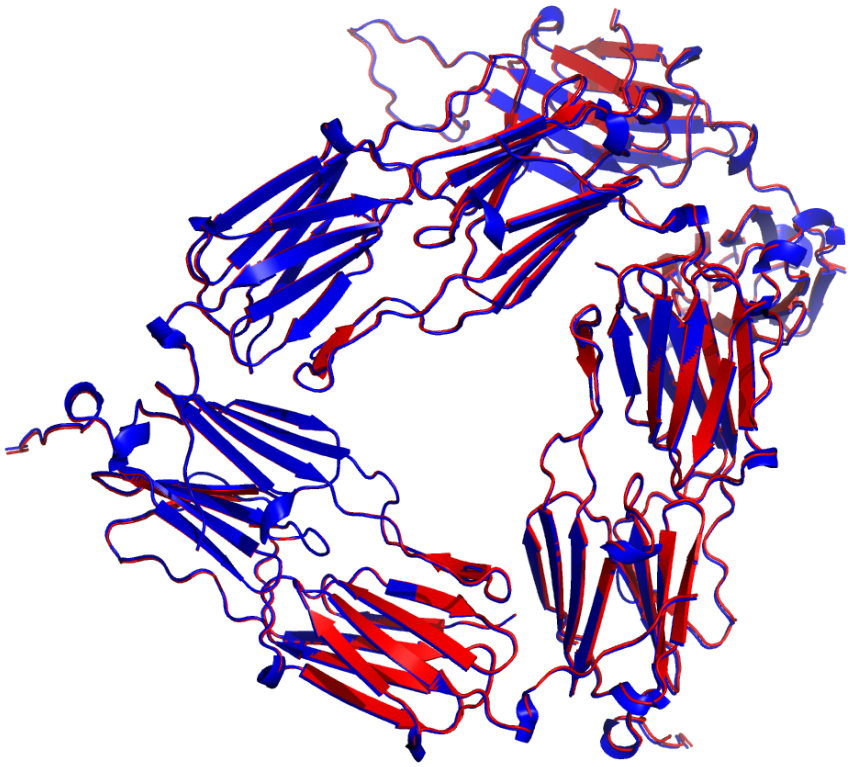


Figure 35: A superimposition of the contents asymmetric unit cells of wild type HSP (blue) and R107G (red), showing the broad similarity between the two structures.

glycine, even at this relatively modest resolution. Statistics from data collection and refinement are included as table 14.

7.6 Summary

For the wild-type protein, the crystal structure has been re-determined at a significantly increased resolution of 2.45Å, representing and increase of 50% in the number of reflections collected. This increased resolution was obtained from crystals grown in the same conditions as those published by the original authors, and most likely represents improvements in purification protocols, coupled with significant improvements in synchrotron X-ray production, diffraction and data

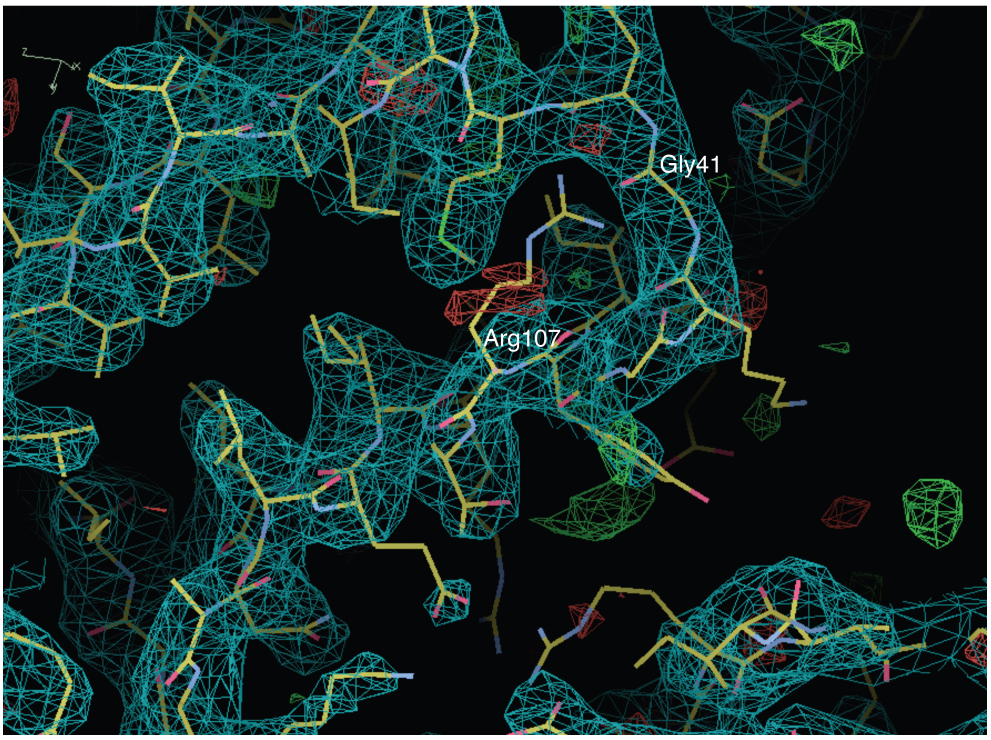


Figure 36: Representative region of the experimental electron density map for R107G contoured at 1.0σ, before refinement. As expected, there is a clear disagreement between the experimental electron density map and the model at position 107, as the model still contained the side chain for arginine.

collection technologies.

For the point mutation R107G, the crystal structure has been determined, and it has been found that this mutation causes no significant effect observable by X-ray crystallography, though differing solution-state behaviours were observed. The two truncation mutations have proved less accessible by X-ray crystallographic techniques, due in part to their greater instability in solution, though the limited data collection that has been possible suggests that this mutation causes a more significant effect in crystallographic terms, though it still does not appear to have significant effect on the quaternary structure of the protein.

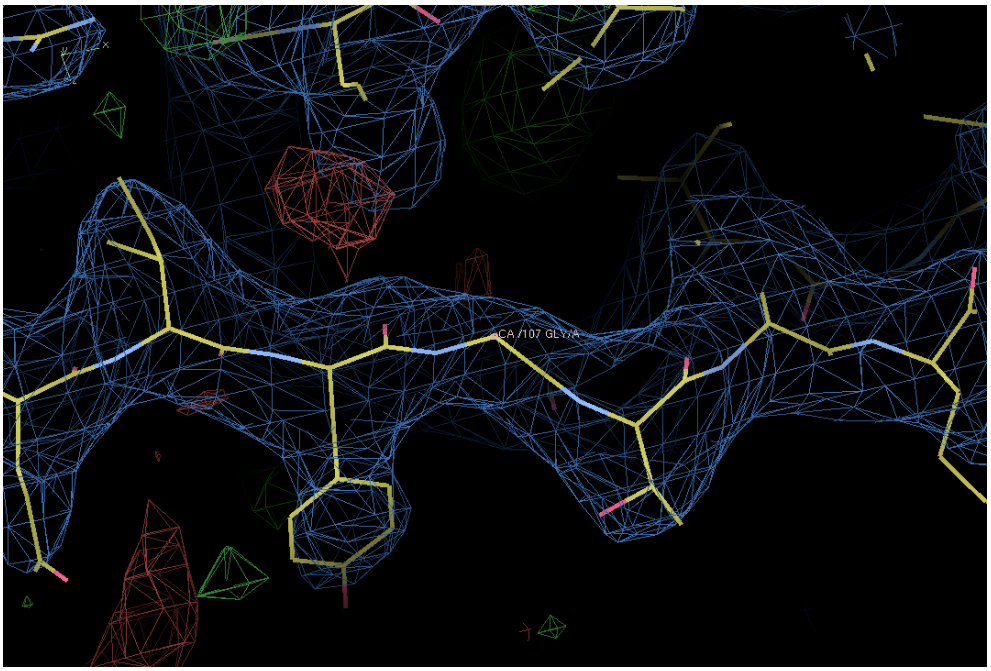


Figure 37: Representative region of the experimental electron density map for R107G contoured at 1.0σ, following refinement with position 107 as a glycine.

Table 14: Data collection and refinement statistics for R107G.

Data Collection Statistics	
Resolution/rA	49.22–2.85
Completeness	99.7%
R_{sym}	0.035
Refinement Statistics	
Unique reflections ($F > 2\sigma$)	26,318
Completeness ($F > 2\sigma$)	97.4%
R factor	0.1803
R_{free}	0.2701
Bond-length deviation/rA	0.011
Bond-angle deviation	1.66°
Mean B Value/Å^2	105.14

8 Conclusions & Future Work

8.1 *Toxoplasma gondii* Serine Palmitoyltransferase

Initial efforts focused on isolating stable quantities of the tagged 181-truncate, as was expected from homology with the other published structures discussed. However, functionality studies demonstrated that this minimal construct did not have serine palmitoyltransferase activity, and a return was therefore made to much longer constructs, in the hope of isolating protein which was both soluble, and functional. It is surprising given the high levels of homology seen with the published structures that this domain was not functional, and indeed it leads one to believe that the acids between 47 and 181 do have some role in the function of the protein. This move to the other extreme, however, brought back the problems which had previously been experienced in trying to obtain significant quantities of constructs with limited solubility, and whilst limited soluble expression was seen in small scale trials for a construct beginning at amino acid 47 with a GST tag from the pGEX system, it was very limited, and proved not to be reproducible on a larger scale.

It is therefore hoped that the results showing significant quantities of soluble expression from a number of the intermediate constructs produced at the Oxford Protein Production Facility may offer a middle ground for moving forwards. These constructs offer such a tantalising step towards crystallisation trials, and thence to, one might hope, diffraction, and a crystal structure, particularly of a construct larger than just the proposed functional domain. This will structure will then finally answer the questions both about the form of the active site and functional unit, and the purpose of the amino acids before acid 181. It is, though, important to remind ourselves of the note of caution that drew us away from the abundant soluble expression of the minimal constructs, and before progressing too rapidly

with crystallisation trials, to ensure that the soluble protein is also functional.

8.1.1 Functionality Studies

The protocol used for testing the functionality of constructs using [^{14}C]-serine is relatively difficult, and expensive to perform, given the involvement of radioactive material. It is also to a large extent qualitative. Moving forward, it would be valuable to be able to use a more simple *in vivo* protocol, which might also offer more quantitative, and conclusive results, particularly given future hopes of producing, and therefore testing, new inhibitors. I therefore suggest the use of the protocol included as appendix H, which avoids the use of radio-isotopes, and instead uses mass spectrometry, giving much more conclusive indications of the presence of the desired species in cell lipid extracts. Using the set of constructs produced at the OPPF this will show at which extent of truncation it is that functionality is lost, again giving further information about the purpose of the domain between amino acids 47 and 181, and informing for which constructs crystallisation ought to be pursued, as it is clear that a crystal structure of a functional protein is far more valuable than one which has no function.

Moving from *in vivo* to *in vitro* Functionality Testing This protocol is, however, still flawed, in that it cannot show for certain that the function of the construct is not reliant upon the host machinery in *E. coli* to in some way supplement its activity. Looking further forward it should therefore be desirable to perform *in vitro* functionality studies. Yard *et al.* use a [^{14}C]-serine-based *in vitro* functionality test [36], and there should be no great difficulty in replacing [^{14}C]-serine with normal serine, and using mass spectrometry instead of TLC and autoradiography in order to detect ketodihydrosphingosine, giving an easy route to *in vitro* functionality testing, where soluble protein is available.

8.1.2 Crystallographic Prospects

The first step in moving forwards to crystallographic studies will be scaling up the 2.5ml high-throughput expression trials to the litre scale. This work should focus on the constructs expressed from pOPINS3C in Rosetta(DE3)pLysS with induction by IPTG. Whilst the most obvious target is the TgSPT144, being the longest construct with very high soluble protein expression, there is certainly some value in also considering the expression of TgSPT90 from this same system. Not only does it also display a good level of expression, but it is also significantly longer, offering the potential to elucidate the nature of the domain between the transmembrane helix and the conserved aminotransferase domain; perhaps that which is most interesting in TgSPT.

8.2 *Methanococcus jannaschii* Small Heat Shock Protein

The redetermination of the structure of HSP16.5 from *M. jannaschii* at higher resolution has laid the ground work for a series of structural studies on mutants designed to mimic particular types of human disease. The first of these to be solved is that of R107G. Whilst there is only limited difference between this structure, and that of the wild type protein, this important result does begin to give us an insight into the relationship between MjHSP16.5 and human αB-crystallin. It leads us to ask the question why a residue with such key effect in αB-crystallin, and which is so highly conserved across all three kingdoms of life, has such limited observable effect observable by crystallographic studies in MjHSP16.5.

The crystal structure clearly shows that the effects of the mutation do not come down to a simple conformational change. It is perhaps possible that it leads only to a change of conformation in solution, an effect which would not be visible in our structure due to crystal lattice constraints. It is also possible that the effect

could come down to a change in dynamic behaviour, which again would not be visible in our crystal structure. The final possibility that I would propose is that, as many other HSPs show many multimeric states in solution, so the R107G mutation could cause changes in this behaviour, something which again would not be visible in the crystal structure.

To further study this mutation, NMR is out of the question for considering the dynamic effects, given the large size of the functional unit. I would therefore suggest that computational methods such as modelling the system as an *elastic network model* may yield results in further understanding this system. For considering the less dynamic solution-state effects, particularly the question of what effect the mutation has on the presence of different multimeric states in solution, I would propose the use of a combination of small angle X-ray scattering and electron microscopy, if the differences in states are sufficient to be distinguished on the electron microscopy grid. Both techniques are able to identify the size of particles, X-ray scattering giving information about the distribution of particle sizes in solution, avoiding the difficulties encountered with the frozen crystallographic state, and electron microscopy being able to give more information about the shape individual discrete states observed. In practice, however, it is only the combination structural and other biophysical techniques which can solve this problem.

The way forward, however, is clear for the truncation mutants. Datasets have been collected for a different set of mutants, 3FR and 4FR, which come tantalisingly close to the 3Å mark. Further improvements in sample quality, coupled with successive improvements in crystallisation conditions are surely all that is necessary to break the resolution barrier, and gain valuable information about how of the C-terminal truncation has affected the surface functionality of MjHSP16.5. This will be particularly important in terms of which residues are unmasked, how

this in turn affects such properties as the hydrophobicity, and how this in turn might affect the protein's properties as a chaperone, and its likeliness to aggregate in cellular conditions. There does, though, still exist the barrier present in the effects of non-crystallographic symmetry on this structure solution, as discussed in section 7.4, which should not be taken lightly. Given the rarity of such systems, the current suite of programs is not well equipped to deal with such situations, and significant work may prove necessary to overcome this hurdle. One possible solution lies in finding the complete assymetric unit contents, and using that as the search model, thus by-passing the difficulties with the current suite of computer programs. This technique was described by Tong in 2001 [134], but given the complexity, and high levels of non-crystallographic symmetry present in this system, it will be by no means straight-forward.

8.3 Overall Summary

Through the auspices of studies into two separate protein systems, the process of protein crystallography has been discussed from cradle to grave. In the first case, the serine palmitoylransfearse from *T. gondii*, we have looked at the early stages of this process, leading up to the isolation of significant amounts of protein in preparation for crystallographic trials. In this we have been successful, demonstrating expression from a number of truncates, and finding that the short, sharp expression of proteins using a SUMO tag and IPTG induction is most successful. We have also gained insights into the domain structure of TgSTP, using both bioinformatic techniques, and *in vivo* isotope labelling studies. We leave this system in a position from which it can immediately move forward to crystallisation trials, and with improved protocols for functionality testing, which offer an easier, quicker, cheaper and more quantitative method for evaluating protein functionality, which will surely prove valuable in any search for active inhibitors.

In the second case we have looked at the small Heat Shock Protein 16.5 from *M. jannaschii*, and three mutants thereof as analogues for disease-causing mutations in human αB-crystallin, starting from the stage of having isolated pure protein, and working through the stages of crystallisation, data collection and structure solution. For the wild-type protein, the crystal structure has been re-determined at a significantly increased resolution of 2.45Å, representing and increase of 50% in the number of reflections collected. This increased resolution was obtained from crystals grown in the same conditions as those published by the original authors, and most likely represents improvements in purification protocols, coupled with significant improvements in synchrotron X-ray production, diffraction and data collection technologies. Whilst this has not extended the knowledge of the protein, it has improved the certainty of our models, and in a system where surface functionality is key, this is valuable knowledge.

For the point mutation R107G, the crystal structure has been determined, and it has been found that this mutation causes no significant effect observable by X-ray crystallography, though differing solution-state behaviours were observed. Whilst this represents the end of crystallographic studies for this mutation, it leaves many questions unanswered, with the door open to further studies into its mode of action. The two truncation mutations have proved less accessible by X-ray crystallographic techniques, due in part to their greater instability in solution, though the limited data collection that has been possible suggests that this mutation causes a more significant effect in crystallographic terms, though it still does not appear to have significant effect on the quaternary structure of the protein. The way forward for these systems is therefore clear, we need to pursue both the quality of crystals, improving the resolution, whilst looking at ways to overcome the difficulties had in solving the structure due to non-crystallographic symmetry, two ends which can be pursued side-by-side.

References

[1] P. Bossart-Whitaker, M. Carson, Y. S. Babu, C. D. Smith, W. G. Laver, and G. M. Air, "Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy 2,3-dehydro-*N*-acetyl neuraminic acid," *Journal of Molecular Biology*, vol. 232, pp. 1069–1083, 1993.

[2] S. Basak and R. W. Compans, "Studies on the role of glycosylation in the functions and antigenic properties of influenza virus glycoproteins," *Virology*, vol. 128, pp. 77–91, 1983.

[3] M. von Itzstein, W.-Y. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. V. Phan, M. L. Smythe, H. F. White, S. W. Oliver, P. M. Colman, J. N. Varghese, D. M. Ryan, J. M. Woods, R. C. Bethell, V. J. Hotham, J. M. Cameron, and C. R. Penn, "Rational design of potent sialidase-based inhibitors of influenza virus replication," *Nature*, vol. 363, pp. 418–423, 1993.

[4] "RCSB PDB Statistics." `http://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html`, February 2012.

[5] W. C. Röntgen, "Über eine neue Art von Strahlen," *Sitzungsberichten der Würzburger Physik.-medic. Geselleschaft*, 1895.

[6] W. Friedrich, P. Knipping, and M. von Laue, "Interferenz-Erscheinungen bei Röntgenstrahlung," *Sitzungsberichte der Mathematisch-Physikalischen Classe der Königlich-Bayerischen Akademie der Wissenschaften zu München*, pp. 303–322, 1912.

[7] W. L. Bragg, "The structure of some crystals as indicated by their diffraction of X-rays," *Proceedings of the Royal Society of London. Series A*, vol. 89, pp. 248–277, 1914.

[8] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, "A three-dimensional model of the myoglobin molecule obtained by

X-ray analysis," *Nature*, vol. 181, pp. 662–666, 1958.

[9] C. R. Groom and F. H. Allen, "Institutional profile: Crystal structure information in drug discovery and development: current perspectives and new possibilities from the cambridge crystallographic data centre.," *Future Medicinal Chemistry*, vol. 2, pp. 933–939, 2010.

[10] P. W. Denny, H. Shams-Eldin, H. P. Price, D. F. Smith, and R. T. Schwarz, "The protozoan inositol phosphorylceramide synthase: A novel drug target that defines a new class of sphingolipid synthase," *Journal of Biological Chemistry*, vol. 218, no. 38, pp. 28200–28209, 2006.

[11] L. M. Weiss and K. Kim, eds., *Toxoplasma gondii. The Model Apicomplexan: Perspectives and Methods*. Academic Press (Elsevier), 2007.

[12] J. P. Dubey and C. P. Beattie, *Toxoplasmosis of Animals and Man*. CRC Press, 1988.

[13] D. Hill and J. P. Dubey, "*Toxoplasma gondii*: transmission, diagnosis and prevention," *Clinical Microbiology & Infection*, vol. 8, pp. 634–640, 2002.

[14] G. Desmonts and J. Couvreur, "Congenital toxoplasmosis. A prospective study of 378 pregnacies.," *New England Journal of Medicine*, vol. 290, pp. 1110–1116, 1974.

[15] B. J. Luft and J. S. Remington, "Toxoplasmic encephalitis in AIDS," *Clinical Infectious Diseases*, vol. 15, pp. 211–222, 1992.

[16] J. J. Aramini, C. Stephen, and J. P. Dubey, "*Toxoplasma gondii* in vancouver island cougars (*Felis concolor vancouverensis*): serology and oocyst shedding.," *Journal of Parasitology*, vol. 84, pp. 438–440, 1998.

[17] J. P. Dubey, K. W. Koula, and A. K. Shara, "Effect of high temperature on infectivity of *Toxoplasma gondii* tissue cysts in pork," *Journal of Food Protection*, vol. 54, pp. 687–690, 1991.

[18] A. W. Kotula, J. P. Dubey, and A. K. Sharar, "Effect of freezing on infectiv-

ity of *Toxoplasma gondii* tissue cysts in pork," *Journal of Food Protection*, vol. 54, pp. 687–690, 1991.

[19] J. P. Dubey and D. W. Thayer, "Killing of different strains of *Toxoplasma gondii* tissue cysts by irradiation under different conditions," *Journal of Parasitology*, vol. 80, pp. 764–767, 1994.

[20] J. L. W. Thudicum, *A Treatise on the Chemical Constitution of the Brain*. Bailliere, Tindall and Cox, London, 1884.

[21] H. E. Carter, F. J. Glick, W. P. Norris, and G. Phillips, "Biochemistry of the sphingolipids. III. Structure of sphingosine," *Journal of Biological Chemistry*, vol. 170, pp. 285–294, 1947.

[22] S. T. Pruett, A. Bushnev, K. Hagedorn, M. Adiga, C. A. Haynes, M. C. Sullards, D. C. Liotta, and A. H. Merrill, "Biodiversity of sphingoid bases ("sphingosines") and related amino acids," *Journal of Lipid Research*, vol. 49, pp. 1621–1639, 2008.

[23] S. Spiegel and S. Milstien, "Sphingosine 1-phosphate, a key cell signaling molecule," *Journal of Biological Chemistry*, vol. 277, no. 29, pp. 25851–25854, 2002.

[24] M. Mayceka, S. Milstien, and S. Spiegel, "Sphingosine-1-phosphate: the swiss army knife of sphingolipid signaling," *Journal of Lipid Research*, vol. 50, pp. S272–S276, 2009.

[25] Y. A. Hannun and L. M. Obeid, "The ceramide-centric universe of lipid-mediated cell regulation: stress encounters of the lipid kind," *Journal of Biological Chemistry*, vol. 29, no. 29, pp. 25847–25870, 2002.

[26] A. H. Futerman and H. Riezman, "The ins and outs of sphingolipid synthesis," *TRENDS in Cell Biology*, vol. 15, pp. 313–318, 2005.

[27] K. Hanada, "Serine palmitoyltransferase, a key enzyme of sphingolipid metabolism," *Biochimica et Biophysica Acta*, vol. 1632, pp. 16–30, 2003.

[28] D. K. Perry, "Serine palmitoyltransferase: role in apoptotic *de novo* ceramide synthesis and other stress responses," *Biochimica et Biophysica Acta*, vol. 1585, 2002.

[29] O. Kerbarh, D. J. Campopiano, and R. L. Baxter, "Mechanism of α-oxoamine synthases: identification of the intermediate claisen product in the 8-amino-7-oxononanoate synthase reaction," *Chemical Communications*, pp. 60–62, 2005.

[30] A. C. Eliot and J. F. Kirsch, "Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations," *Annual Review of Biochemistry*, vol. 73, pp. 383–415, 2004.

[31] D. Alexeev, M. Alexeeva, R. L. Baxter, D. J. Campopiano, S. P. Webster, and L. Sawyer, "The crystal structure of 8-amino-7-oxononanoate synthase: A bacterial PLP-dependent, acyl-CoA-condensing enzyme," *Journal of Molecular Biology*, vol. 284, pp. 401–419, 1998.

[32] S. P. Webster, D. Alexeev, D. J. Campopiano, R. M. Watt, M. Alexeeva, L. Sawyer, and R. L. Baxter, "Mechanism of 8-amino-7-oxononanoate synthase: Spectroscopic, kinetic, and crystallographic studies," *Biochemistry*, vol. 39, no. 3, pp. 516–528, 2000.

[33] I. Astner, J. O. Schulze, J. van den Heuvel, D. Jahn, W.-D. Schubert, and D. W. Heinz, "Crystal structure of 5-aminolevulinate synthase the first enzyme of heme biosynthesis and its link to XLSA in humans," *The EMBO Journal*, vol. 24, pp. 3166–3177, 2005.

[34] Z. Zaman, P. M. Jordan, and M. Akhtar, "Mechanism and stereochemistry of the 5-aminolaevulinate synthetase reaction," *Biochemistry Journal*, vol. 135, pp. 257–263, 1973.

[35] A. Schmidt, J. Sivaraman, Y. Li, R. Larocque, J. A. R. G. Barbosa, C. Smith, A. Matte, J. D. Schrag, and M. Cygler, "Three-dimensional struc-

ture of 2-amino-3-ketobutyrate CoA ligase from *Escherichia coli* complexed with a PLPsubstrate intermediate: Inferred reaction mechanism," *Biochemistry*, vol. 40, no. 17, pp. 5151–5160, 2001.

[36] B. A. Yard, L. G. Carter, K. A. Johnson, I. M. Overton, M. Dorward, H. Liu, S. A. McMahon, M. Oke, D. Puech, G. J. Barton, J. H. Naismith, and D. J. Campopiano, "The structure of serine palmitoyltransferase; gateway to sphingolipid biosynthesis," *Journal of Molecular Biology*, vol. 370, pp. 870–886, 2007.

[37] P. W. Denny, S. Pratt, C. Bruce, and C. J. Barnes *Unpublished work.*

[38] C. J. Barnes, "Sphingolipids in *Toxoplasma gondii*, synthesis and scavenging," Master's thesis, University of Durham, 2010.

[39] R. Buede, C. Rinker-Schaffer, W. J. Pinto, R. L. Lester, and R. C. Dickson, "Cloning and characterization of *lcb1*, a *Saccharomyces* gene required for biosynthesis of the long-chain base component of sphingolipids," *Journal of Bacteriology*, vol. 173, no. 14, pp. 4325–4332, 1991.

[40] G. Han, K. Gable, L. Yan, M. Natarajan, J. Krishnamurthy, S. D. Gupta, A. Borovitskaya, J. M. Harmon, and T. M. Dunn, "The topology of the LCB1p subunit of yeast serine palmitoyltransferase," *Journal of Biochemistry*, vol. 279, pp. 53707–53716, 2004.

[41] H. Ikushiro, H. Hayashi, and H. Kagamiyama, "A water-soluble homodimeric serine palmitoyltransferase from *Sphingomonas paucimobilis* EY2395^T strain," *Journal of Biochemistry*, vol. 276, pp. 18249–18256, 2001.

[42] H. Ikushiro, M. M. Islam, H. Tojo, and H. Hayashi, "Molecular characterization of membrane-associated soluble serine palmitoyltransferases from *Sphingobacterium multivorum* and *Bdellovibrio stolpii*," *Journal of Bacteriology*, vol. 189, pp. 5749–5761, 2007.

[43] M. M. Nagiec, R. L. Lester, and R. C. Dickson, "Sphingolipid synthe-

sis: identification and characterization of mammalian cDNAs encoding the LCB2 subunit of serine palmitoyltransferase," *Gene*, vol. 177, pp. 237–241, 1996.

[44] K. Hanada, T. Hara, M. Nishijima, O. Kuge, R. C. Dickson, and M. M. Nagiec, "A mammalian homolog of the yeast *lcb1* encodes a component of serine palmitoyltransferase, the enzyme catalyzing the first step in sphingolipid synthesis," *Journal of Biochemistry*, vol. 272, pp. 32108–32114, 1997.

[45] G. Schneider, Helena Käck, and Y. Lindqvist, "The manifold of vitamin B_6 dependent enzymes," *Structure*, vol. 8, pp. R1–R6, 2000.

[46] A. Mozzarelli and S. Bettati, "Exploring the pyridoxal 5-phosphate-dependent enzymes," *The Chemical Record*, vol. 6, pp. 275–287, 2006.

[47] E. A. Peterson and H. A. Sober, "Preparation of crystalline phosphorylated derivatives of vitamin B_6," *Journal of the American Chemical Society*, vol. 76, pp. 169–175, 1954.

[48] S. D. Lvova, A. L. Wkovleva, I. S. Cherkasova, M. V. Balyakina, and V. I. Gunar, "Method of purification and separtion of pyridoxal-5'-phosphate," *Khimiko-Farmatsevticheskii Zhurnal*, vol. 9, no. 11, pp. 23–25, 1975.

[49] G. A. Hunter and G. C. Ferreira, "Lysine-313 of 5-aminolevulinate synthase acts as a general base during formation of the quinonoid reaction intermediates," *Biochemistry*, vol. 38, no. 12, pp. 3711–3718, 1999.

[50] P. K. Mehta and P. Christen, "The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes," *Advances in Enzymology and Related Areas of Molecular Biology*, vol. 74, pp. 129–189, 2000.

[51] A. Amadasi, M. Bertoldi, R. Contestabile, S. Bettati, B. Cellini, M. L. di Salvo, C. Borri-Voltattorni, F. Bossa, and A. Mozzarelli, "Pyridoxal 5'-phosphate enzymes as targets for therapeutic agents," *Current Medicinal*

Chemistry, vol. 14, pp. 1291–1324, 2007.

[52] K. A. Medlock and A. H. Merrill Jr., "Inhibition of serine palmitoyltransferase *in vitro* and long-chain base biosynthesis in intact Chinese hamster ovary cells by β-chloroalanine," *Biochemistry*, vol. 27, pp. 7079–7084, 1988.

[53] K. S. Sundaram and M. Lev, "Inhibition of sphingolipid synthesis by cycloserine *in vitro* and *in vivo*," *Journal of Neurochemistry*, vol. 42, pp. 577–581, 1984.

[54] Y. Miyake, Y. Kozutsumi, S. Nakamura, T. Fujita, and T. Kawasaki, "Serine palmitoyltransferase is the primary target of a sphingosine-like immunosuppressant, ISP-1/myriocin," *Biochemical and Biophysical Research Communications*, vol. 211, pp. 396–403, 1995.

[55] Y. Sun, R. Taniguchi, D. Tanoue, T. Yamaji, H. Takematsu, K. Mori, T. Fujita, T. Kawasaki, and Y. Kozutsumi, "Sli2 (Ypk1), a homologue of mammalian protein kinase SGK, is a downstream kinase in the sphingolipid-mediated signaling pathway of yeast," *Molecular and Cell Biology*, vol. 20, pp. 4411–4419, 2000.

[56] M. Momoi, D. Tanoue, Y. Sun, H. Takematsu, Y. Suzuki, M. Suzuki, A. Suzuki, T. Fujita, and Y. Kozutsumi, "SLi1 (YGR212W) is a major gene conferring resistance to sphingolipid biosynthesis inhibitor ISP-1, and encodes an ISP-1 *N*-acetyltransferase in yeast," *Biochemical Journal*, vol. 381, pp. 321–328, 2004.

[57] A. Yamaji-Hasegawa, A. Takahashi, Y. Tetsuka, Y. Senoh, and T. Kobayashi, "Fungal metabolite sulfamisterin suppresses sphingolipid synthesis through inhibition of serine palmitoyltransferase," *Biochemistry*, vol. 44, pp. 268–277, 2005.

[58] F. Ritossa, "A new puffing pattern induced by temperature shock and DNP in *Drosophila*," *Experientia*, vol. 18, pp. 571–573, 1962.

[59] M. Haslbeck, T. Franzmann, D. Weinfurtner, and J. Buchner, "Some like it hot: the structure and function of small heat-shock proteins," *Nature Structural & Molecular Biology*, vol. 12, no. 10, pp. 842–846, 2005.

[60] P. Laksanalamai and F. T. Robb, "Small heat shock proteins from extremophiles: a review," *Extremophiles*, vol. 8, pp. 1–11, 2004.

[61] Y. Sun and T. H. MacRae, "Small heat shock proteins: molecular structure and chaperone function," *Cellular and Molecular Life Sciences*, vol. 62, pp. 2460–2476, 2005.

[62] C. Eifert, M. R. Burgio, P. M. Bennett, J. C. Salernoa, and J. F. Koretz, "N-terminal control of small heat shock protein oligomerization: Changes in aggregate size and chaperone-like function," *Biochimica et Biophysica Acta*, vol. 1748, pp. 146–156, 2005.

[63] P. Thampi and E. C. Abraham, "Influence of the C-terminal residues on oligomerization of αA-crystallin," *Biochemistry*, vol. 42, pp. 11857–11863, 2003.

[64] P. Fernando, R. Abdulle, A. Mohindra, J. G. Guillemette, and J. J. Heikkila, "Mutation or deletion of the C-terminal tail affects the function and structure of *Xenopus laevis* small heat shock protein, hsp30," *Comparative Biochemistry and Physiology - Part B: Biochemistry & Molecular Biology*, vol. 133, pp. 95–103, 2002.

[65] J. H. Liao, J. S. Lee, and S. H. Chiou, "C-terminal lysine truncation increases thermostability and enhances chaperone-like function of porcine αB-crystallin," *Biochemical and Biophysical Research Communications*, vol. 297, pp. 309–316, 2002.

[66] H. Bloemendal and W. W. de Jong, "Lens proteins and their genes," *Progress in Nucleic Acid Research and Molecular Biology*, vol. 41, pp. 259–281, 1991.

[67] R. Klemenz, E. Fröhli, R. H. Steiger, R. Schäfer, and A. Aoyama, "αb-crystallin is a small heat-shock protein," *Proceedings of the National Academy of Science*, vol. 88, pp. 3652–3656, 1991.

[68] R. A. Dubin, E. F. Wawrousek, and J. Piatigorsky, "Expression of the marine αb-crystallin gene is not restricted to the lens," *Molecular and Cell Biology*, vol. 9, pp. 1083–1091, 1989.

[69] S. P. Bhat and C. N. Nagineni, "αb subunit of lens specificprotein α-crystallin is present in other ocular and nonocular tissues," *Biochemical and Biophysical Research Communications*, vol. 158, pp. 319–325, 1989.

[70] P. Vicart, A. Caron, P. Guicheney, Z. Li, M. C. Prevost, A. Faure, D. Chateau, F. Chapon, F. Tome, J. M. Dupret, D. Paulin, and M. Fardeau, "A missense mutation in theαb-crystallin chaperone gene causes a desmin-related myopathy," *Nature Genetics*, vol. 20, pp. 92–95, 1998.

[71] V. Berry, P. Francis, M. A. Reddy, D. Collyer, E. Vithana, I. MacKay, G. Dawson, A. H. carey, A. Moore, S. S. Bhattacharya, and R. A. Quinlan, "α-b crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans," *American Journal of Human Genetics*, vol. 69, pp. 1141–1145, 2001.

[72] D. Selcen and A. G. Engel, "Myofibrillar myopathy caused by novel dominant negative αB-crystallin mutations," *Annals of Neurology*, vol. 54, pp. 804–810, 2003 2003.

[73] V. H. Hayes, G. Devlin, and R. A. Quinlan, "Truncation of α-B-crystallin by the myopathy-causing Q151X mutation significantly destabilizes the protein leading to aggregate formation in transfected cells," *Journal of Biological Chemistry*, vol. 283, pp. 10500 –10512, 2008.

[74] C. Bagnéris, O. A. Bateman, C. E. Naylor, N. Cronin, W. C. Boelens, N. H. Keep, and C. Slingsby, "Crystal structures of α-crystallin domain dimers of

αB-crystallin and hsp20," *Journal of Molecular Biology*, vol. 392, pp. 1242–1252, 2009.

[75] D. A. Haley, M. P. Bova, Q.-L. Huang, H. S. Mchaourab, and P. L. Stewart, "Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies," *Journal of Molecular Biology*, vol. 298, pp. 261–272, 2000.

[76] J. Horwitz, M. P. Bova, L.-L. Ding, D. A. Haley, and P. L. Stewart, "Lens α-crystallin: Function and structure," *Eye*, vol. 13, pp. 403–408, 1999.

[77] D. A. Haley, J. Horwitz, and P. L. Stewart, "The small heat-shock protein, αb-crystallin, has a variable quaternary structure," *Journal of Molecular Biology*, vol. 277, pp. 27–35, 1998.

[78] W. J. Jones, J. A. Leigh, F. Mayer, C. Woese, and R. S. Wolfe, "*Methanococcus jannaschii sp-nov* , an extremely thermophiliic methanogen from a submarine hydrothermal vent," *Archives of Microbiology*, vol. 136, pp. 254–261, 1983.

[79] A. Elofsson and E. L. Sonnhammer, "A comparison of sequence and structure protein domain families as a basis for structural genomics.," *Bioinformatics*, vol. 15, pp. 480–500, 1999.

[80] T. I. Zarembinski, L. W. Hung, H. J. Mueller-Dieckmann, K. K. Kim, H. Yokota, R. Kim, and S. H. Kim, "Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics.," *Proceedings of the National Academy of Science*, vol. 95, pp. 15189–15193, 1998.

[81] K. K. Kim, H. Yokota, S. Santoso, D. Lerner, R. Kim, and S.-H. Kim, "Purification, crystallization, and preliminary X-ray crystallographic data analysis of small heat shock protein homolog from *Methanococcus jannaschii*, a hyperthermophile," *Journal of Structural Biology*, vol. 121, pp. 76–80, 1998.

[82] K. K. Kim, R. Kim, and S.-H. Kim, "Crystal structure of a small heat-shock protein," *Nature*, vol. 394, pp. 595–599, 1998.

[83] R. Kim, K. K. Kim, H. Yokota, and S.-H. Kim, "Small heat shock protein of *Methanococcus jannaschii*, a hyperthermophile," *Proceedings of the National Academy of Science*, vol. 95, pp. 9129–9133, 1998.

[84] R. Kim, L. Lai, H.-H. Lee, G.-W. Cheong, K. K. Kim, Z. Wu, H. Yokota, S. Marqusee, and S.-H. Kim, "On the mechanism of chaperone activity of the small heat-shock protein of *Methanococcus jannaschii*," *Proceedings of the National Academy of Science*, vol. 100, pp. 8151–8155, 2003.

[85] D. R. Kim, I. Lee, S. C. Ha, and K. K. Kim, "Activation mechanism of HSP16.5 from *Methanococcus jannaschii*," *Biochemical and Biophysical Research Communications*, vol. 307, pp. 991–998, 2003.

[86] Z. Wang, A. Cao, and L. Lai, "High activity of Mj HSP16.5 under acidic condition," *Science in China Series B: Chemistry*, vol. 52, pp. 325–331, 2009.

[87] J. Shi, Hanane, Koteiche, H. S. Mchaourab, and P. L. Stewart, "Cryoelectron microscopy and EPR analysis of engineered symmetric and polydisperse HSP16.5 assemblies reveals determinants of polydispersity and substrate binding," *Journal of Biological Chemistry*, vol. 281, pp. 40420–40428, 2006.

[88] A. Cao, Z. Wang, P. Wei, F. Xu, J. Cao, and L. Lai, "Preheating induced homogeneity of the small heat shock protein from *Methanococcus jannaschii*," *BIiochimica et Biophysica Acta*, vol. 1748, pp. 489–495, 2008.

[89] Y. Guan, Z. Wang, A. Cao, L. Lai, and X. S. Zhao, "Subunit exchange of MjHSP16.5 studied by single-molecule imaging and fluorescence resonance energy transfer," *Journal of the American Chemical Society*, vol. 128, no. 22, pp. 7203–7208, 2006.

[90] B. Rupp, *Biomolecular Crystallography: Principles, Practice, and Applica-*

tion to Structu Biology. Garland Science (Taylor & Francis Group), 2010.

[91] S. Tabor and C. C. Richardson, "A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes," *Proceedings of the National Academy of Science*, vol. 82, pp. 1074–1078, 1985.

[92] T. G. M. Schmidt and A. Skerra, "The Strep-tag system for one-step purification and high-affinity detection or capturing of protein," *Nature Protocols*, vol. 2, pp. 1528–1532, 2007.

[93] R. B. Kapust and D. S. Waugh, "*Escherichia coli*maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused," *Protein Science*, vol. 8, pp. 1668–1674, 1999.

[94] W. Kaplan, P. Hüsler, H. Klump, J. Erhardt, N. Sluis-Cremer, and H. Dirr, "Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag," *Protein Science*, vol. 6, pp. 399–406, 1997.

[95] "Blast: Basic local alignment search tool," February 2012. `http://blast.ncbi.nlm.nih.gov/`.

[96] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, pp. 3389–3402, 1997.

[97] "Jpred 3: A secondary structure prediction server," February 2012. `http://www.compbio.dundee.ac.uk/www-jpred/`.

[98] C. Cole, J. D. Barber, and G. Barton, "The Jpred 3 secondary structure prediction server," *Nucleic Acids Research*, vol. 35, pp. W197–W201, 2008.

[99] "Hmmtop: Predic of transmembrane helices and topology of proteins," February 2012. `http://www.enzim.hu/hmmtop/`.

[100] G. E. Tusnády and I. Simon, "The HMMTOP transmembrane topology prediction server," *Bioinformatics*, vol. 17, pp. 849–850, 2001.

[101] G. E. Tusnády and I. Simon, "Principles governing amino acid composition of integral membrane proteins: Applications to topology prediction," *Journal of Molecular Biology*, vol. 283, pp. 489–506, 1998.

[102] K. B. Mullis and F. A. Faloona, "Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction," *Methods in Enzymology*, vol. 155, pp. 335–350, 1987.

[103] C. Aslanidis and P. J. de Jong, "Ligation-independent cloning of PCR products (LIC-PCR)," *Nucleic Acids Research*, vol. 18, pp. 6069–6074, 1990.

[104] F. W. Studier, "Protein production by auto-induction in high-density shaking cultures," *Protein Expression and Purification*, vol. 41, pp. 207–234, 2005.

[105] J. Phan, A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters 3rd, R. B. Kapust, M. Li, A. Wlodawer, and D. S. Waugh, "Structural basis for the substrate specificity of tobacco etch virus protease," *Journal of Biological Chemistry*, vol. 277, pp. 50564–50572, 2002.

[106] T. D. Parks, K. K. Leuther, eric D Howard, S. A. Johnston, and W. G. Dougherty, "Release of proteins and peptides from fusion proteins using a recombinant plant virus protease," *Analytical Biochemistry*, vol. 216, pp. 413–417, 1994.

[107] M. G. Cordingley, R. B. Register, P. L. Callahan, V. M. Garsky, and R. J. Colonno, "Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in *Escherichia coli*.," *Journal of Virology*, vol. 63, pp. 5037–5045, 1989.

[108] T. M. Bergfors, ed., *Protein Crystallisation. Techniques, Strategies and Tips. A Laborartory Manual*. International University Line, 1999.

[109] J. W. Wooh, R. D. Kidd, J. L. Martin, and B. Kobe, "Comparison of three commercial sparse-matrix crystallization screens.," *Acta Crystallographica, Section D: Biological Crystallography*, vol. 59, pp. 769–772, 2003.

[110] E. F. Garman, "Radiation damage in macromolecular crystallography: what is it and why should we care?," *Acta Crystallographica, Section D: Biological Crystallography*, vol. D66, pp. 339–351, 2010.

[111] E. F. Garman and E. P. Mitchell, "Glycerol concentrations required for cryoprotection of 50 typical protein crystallisation solutions," *Journal of Applied Crystallography*, vol. 29, pp. 584–587, 1996.

[112] T. Hahn, ed., *International Tables for Crystallography, Volume A: Space-group symmetry*. IUCr, 2006.

[113] G. L. Taylor, "Introduction to phasing.," *Acta Crystallographica, Section D: Biological Crystallography*, vol. 66, pp. 325–338, 2010.

[114] M. G. Rossmann, "Molecular replacement–historical background.," *Acta Crystallographica, Section D: Biological Crystallography*, vol. 57, pp. 1360–1366, 2001.

[115] M. G. Rossmann, "The molecular replacement method.," *Acta Crystallographica, Section A: Foundations of Crystallography*, vol. 46, pp. 73–82, 19990.

[116] W. A. Hendrickson, "Determination of macromolecular structure from anomalous diffraction of synchrotron radiation," *Science*, vol. 254, pp. 51–58, 1991.

[117] R. W. James, *The Optical Principles of the Diffraction of X-rays*. Bell, London, 1948.

[118] "RCSB PDB." `http://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html`, February 2012.

[119] "Ebi clustalw - multiple sequence alignment." `http://www.ebi.ac.uk/`

Tools/msa/clustalw2/, February 2012.

[120] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins, "ClustalW and ClustalX version 2," *Bioinformatics*, vol. 23(21), pp. 2947–2948, 2007.

[121] M. Goujon, H. M. W. Li, F. Valentin, S. Squizzato, J. Paern, and R. Lopez, "A new bioinformatics analysis tools framework at EMBL-EBI," *Nucleic Acids Research*, vol. 38 Suppl., pp. W695–W699, 2010.

[122] "ExPASy ProtParam." http://web.expasy.org/protparam/.

[123] J. M. Walker, ed., *The Proteomics Protocols Handbook*. Humana Press, 2005.

[124] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, and A. Sali, "Comparative protein structure modeling using MODELLER," *Current Protocols in Bioinformatics*, vol. Chapter 5, p. Unit 5.6, 2006.

[125] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of *coot*," *Acta Crystallographica, Section D: Biological Crystallography*, vol. D66, pp. 486–501, 2010.

[126] J. Jancarik, R. Pufan, C. Hong, S. H. Kim, and R. Kim, "Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins," *Acta Crystallographica, Section D: Biological Crystallography*, vol. D60, pp. 1670–1673, 2004.

[127] Collaborative Computational Project, Number 4, "The CCP4 suite: Programs for protein crystallography," *Acta Crystallographica, Section D: Biological Crystallography*, vol. D50, pp. 760–763, 1994.

[128] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method," *Acta Crystallo-*

graphica, Section D: Biological Crystallography, vol. D53, pp. 240–255, 1997.

[129] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics," *Acta Crystallographica, Section D: Biological Crystallography*, vol. D60, pp. 2126–2132, 2004.

[130] H. Ikushiro, M. M. Islam, A. Okamoto, J. Hoseki, T. Murakawa, S. Fujii, I. Miyahara, and H. Hayashi, "Insights into the enzymayic mechanism of serine palmitoyltransferase from *sphingobacterium multivorum*," *Journal of Biochemistry*, vol. 146, p. 549, 2009.

[131] A. Sali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Journal of Molecular Biology*, vol. 234, pp. 779–815, 1993.

[132] A. Fiser, R. K. Do, and A. Sali, "Modeling of loops in protein structures," *Protein Science*, vol. 9, pp. 1753–1773, 2000.

[133] M. G. Rossmann and D. M. Blow, "The detection of sub-units within the crystallographic asymmetric unit," *Acta Crystallographica*, vol. 15, pp. 24–31, 1962.

[134] L. Tong, "How to take advantage of non-crystallographic symmetry in molecular replacement: 'locked' rotation and translation functions.," *Acta Crystallographica, Section D: Biological Crystallography*, vol. 57, pp. 1381–1389, 2001.

A Serine Palmitoyltransferase Sequences

A.1 Full length protein:

MASGATYFTRGTGSPFLGAGVEWASNIDLFLCAFLSASVLGILLAFFNDEVSWGSLRWSW
IATQLLPITPCSSHAVYKDVETALAKAARNKAGSKRALEEFLAALQDGTVMVLLSKWSAR
GFERLAFYWQALKIKYTAQSRRQFFYQMQKVQLKLEIKPGETEMQSYNDAKRYMKSRDLW
PFAYEVSNVKDTQVICEGVRAYPMSSYSYLDFVREPLQEAALAAGRTWSTGNHGARMLGG
NMRILRDLEKMVGRFFGREDSLLCATGFLATMSSICAVAKEGDLIVGDNRLHASLRSGMK
LSGAKEMLFRHNNWHHLQQTLAKHRRKYKNCWIVIESVYSMDGDIADLPVVRRLADQYNC
RILLDEAHGLGVLGKTGRGLEEHFNMPGAADVIVGTFSKSIGGVGGYITGDNDLVEFLDF
HAPGSVFSAPLTAYSAGGAMMAFELMQGEQSWRIAKAQENAKYLRHALQTGLGLWPKDYP
AERKFELEGVACTTVIPVVFPHDGDRVFRVTQAMLKRGWMVAAAAYPACPLNRPRIRVTA
TAAYNQKMMDEFVKSLVEVTVECPPTDMLR

A.2 Fusion protein 181 from pETzt2_1

MKHHHHHHPMKQHDEAVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANL
LAEAKKLNDAQAPKVDAGSGSGSENLYFQGAMAPFAYEVSNVKDTQVICEGVRAYPMSSY
SYLDFVREPLVQEAALAAGRTWSTGNHGARMLGGNMRILRDLEKMVGRFFGREDSLLCAT
GFLATMSSICAVAKEGDLIVGDNRLHASLRSGMKLSGAKEMLFRHNNWHHLQQTLAKHRR
KYKNCWIVIESVYSMDGDIADLPVVRRLADQYNCRILLDEAHGLGVLGKTGRGLEEHFNM
PGAADVIVGTFSKSIGGVGGYITGDNDLVEFLDFHAPGSVFSAPLTAYSAGGAMMAFELM
QGEQSWRIAKAQENAKYLRHALQTGLGLWPKDYPAERKFELEGVACTTVIPVVFPHDGDR
VFRVTQAMLKRGWMVAAAAYPACPLNRPRIRVTATAAYNQKMMDEFVKSLVEVTVECPPT
DMLR

Section	Colour
His tag	Green
ZTag2	Blue
TEV Recognition Site	Red
181 Truncate	Orange

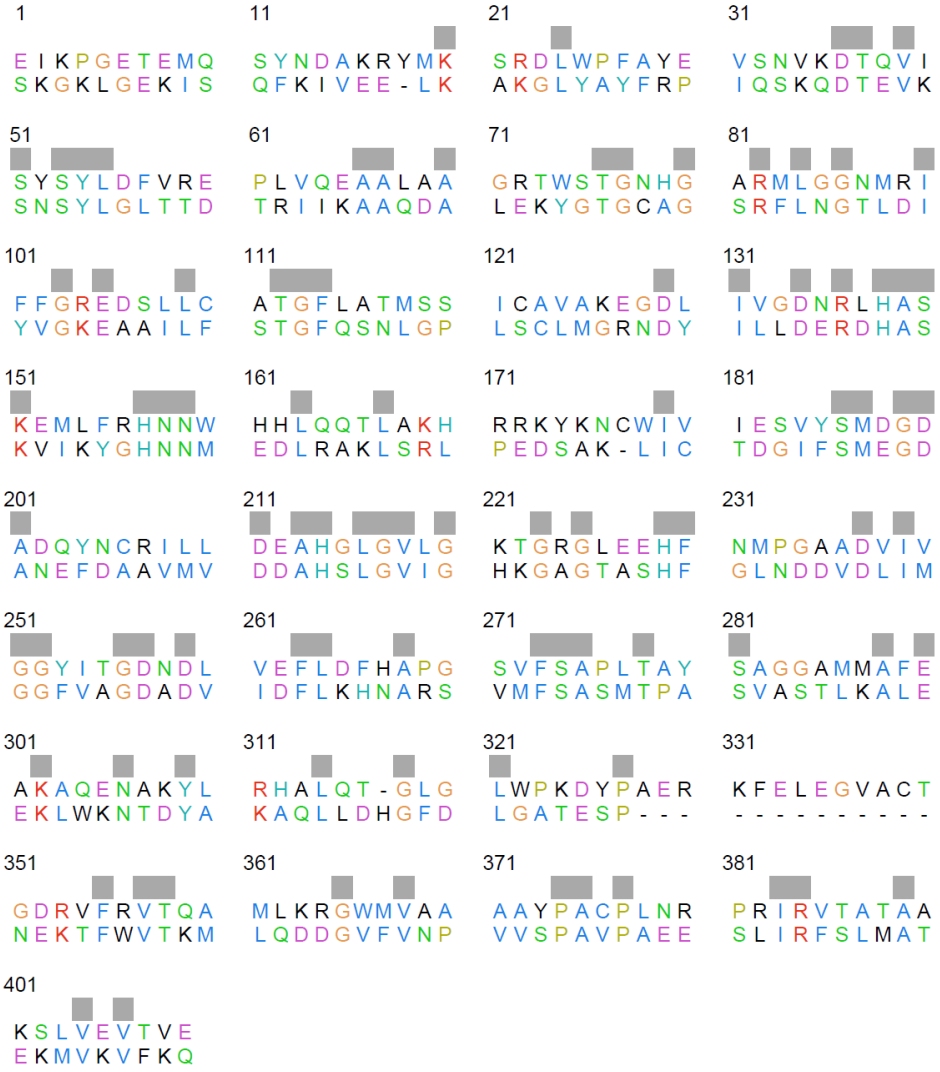
B Serine Palmitoyltransferase Alignment

```
SpSPT          -AIVMEQVKSPTEAVIRGKDTILLGTYNYMGMTFDPDVIAAGKEALEKFGSGTNGSRMLN 106
EcAKCL         ---IITSAQQADITVADGSHVINFCANNYLGLANHPDLIAAAKAGXDSHGFGXASVRFIC 86
RcALS          ----AFPKAQWNRPDGGKQDITVWCGNDYLGMGQHPVVLAAMHEALEAVGAGSGGTRNIS 87
TgSPT1         PFAYEVSNVKDTQVICEGVRAYPMSSYSYLDFVREPLQEAALAAG-RTWSTGNHGARMLG 239
                         .                  .*:.:  .*   **   .    . *  . * :

SpSPT          GTFHDHMEVEQALRDFYGTTGAIVFSTGYMANLGIISTLAGKGE--YVILDADSHASIYD 164
EcAKCL         GTQDSHKELEQKLAAFLGXEDAILYSSCFDANGGLFETLLGAED--AIISDALNHASIID 144
RcALS          GTTAYHRRLEAEIAGLHQKEAALVFSSAYNANDATLSTLRVLFPGLIIYSDSLNHASMIE 147
TgSPT1         GNMRILRDLEKMVGRFFGREDSLLCATGFLATMSSICAVAKEGD--LIVGDNRLHASLRS 297
               *.      :*  :  :     ::: :: : *. . : ::        :  *   ***: .

SpSPT          GCQQGNAEIVRFRHNSVEDLDKRLGRLPKEPAK--LVVLEGVYSMLGDIAPLKEMVAVAK 222
EcAKCL         GVRLCKAKRYRYANNDXQELEARLKEAREAGARHVLIATDGVFSXDGVIANLKGVCDLAD 204
RcALS          GIKRNAGPKRIFRHNDVAHLRELIAADDPAAPK--LIAFESVYSMDGDFGPIKEICDIAE 205
TgSPT1         GMKLSGAKEMLFRHNNWHHLQQTLAKHRRKYKN-CWIVIESVYSMDGDIADLPVVRRLAD 356
               * :   .    : :*.  .*   :        .   :. :.*:*  * :. :  :  :*.

SpSPT          KHGAMVLVDEAHSMGFFGPNGRGVYEAQGLEGQIDFVVGTFSKSVG-TVGGFVVSNHPKF 281
EcAKCL         KYDALVXVDDSHAVGFVGENGRGSHEYCDVXGRVDIITGTLGKALGGASGGYTAARKEVV 264
RcALS          EFGALTYIDEVHAVGMYGPRGAGVAERDGLMHRIDIFNGTLAKAYG-VFGGYIAASARMV 264
TgSPT1         QYNCRILLDEAHGLGVLGKTGRGLEEHFNMPGAADVIVGTFSKSIG-GVGGYITGDNDLV 415
               :...   :*: *.:*. *  * *  *  .:    *.. **:.*: *   **: ..    .

SpSPT          EAVRLACRPYIFTASLPPSVVATATTSIRKLMTA--HEKRERLWSNARALHGGLKAMG-- 337
EcAKCL         EWLRQRSRPYLFSNSLAPAIVAASIKVLEXVEAG--SELRDRLWANARQFREQXSAAG-- 320
RcALS          DAVRSYAPGFIFSTSLPPAIAAGAQASIAFLKTAEGQKLRDAQQMHAKVLKMRLKALG-- 322
TgSPT1         EFLDFHAPGSVFSAPLTAYSAGGAMMAFELMQGEQ-SWRIAKAQENAKYLRHALQTGLGL 474
               : :   .   :*: .*..  .. :   :  :              :*: ::   .:

SpSPT          ----------FRLGTETCDSAIVAVMLEDQEQAAMMWQALL-DGGLYVNMARPPATPAGT 386
EcAKCL         ----------FTL--AGADHAIIPVXLGDAVVAQKFARELQ-KEGIYVTGFFYPVVPKGQ 367
RcALS          ----------MPI--IDHGSHIVPVVIGDPVHTKAVSDMLLSDYGVYVQPINFPTVPRGT 370
TgSPT1         WPKDYPAERKFELEGVACTTVIPVVFPHDGDRVFRVTQAML-KRGWMVAAAAYPACPLNR 533
                         : :        *  *   *   .  .   :  . *  *     *. * .

SpSPT          FLLRCSICAEHTPAQIQTVLGMFQAAGRAVGVIGLE- 422
EcAKCL         ARIRTQXSAAHTPEQITRAVEAFTRIGKQLGVIA--- 401
RcALS          ERLRFTPSPVHDLKQIDGLVHAMDLLWARCA------ 401
TgSPT1         PRIRVTATAAYNQKMMDEFVKSLVEVTVECPPTDMLR 570
                 :*    . :    :   :  :
```

Abbreviation	Enzyme
SpSPT	*Sphingomonas paucimobilis* Serine Palmitoyltransferase
EcAKCL	*Escherichia coli* 2-Amino-3-ketobutyrate CoA Ligase
RcALS	*Rhodobacter capsulatis* 5-Aminolaevulinate Synthetase
TgSPT	*Toxoplasma gondii* Serine Palmitoyltransferase

Amino Acids	Colour	Characteristics
AVFPMILW	Red	Small (Small & Hydrophobic (Incl. Aromatic -Y))
DE	Blue	Acidic
RK	Magenta	Basic
STYHCNGQ	Green	Hydroxyl + Amine + Basic - Q

C Serine Palmitoyltransferase Homology Model Alignment



This sequence alignment was performed with ClustalX using the standard colour-coding.

D Serine Palmitoyltransferase Secondary Structure Prediction

```
MASGATYFTRGTGSPFLGAGVEWASNIDLFLCAFLSASVLGILLAFFNDEVSWGSLRWSW
CCCCCEEEECCCCCCHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCHCHHH

IATQLLPITPCSSHAVYKDVETALAKAARNKAGSKRALEEFLAALQDGTVMVLLSKWSAR
HHHHHCCCCCCCHHHHCHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHCCCHHHHHHHHHH

GFERLAFYWQALKIKYTAQSRRQFFYQMQKVQLKLEIKPGETEMQSYNDAKRYMKSRDLW
HHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHCCCC

PFAYEVSNVKDTQVICEGVRAYPMSSYSYLDFVREPLVQEAALAAGRTWSTGNHGARMLG
CCCEEECCCCCCEEEECCCEEEECCCCCHCCCCCCHHHHHHHHHHHHHHCCCCCHHHHHC

GNMRILRDLEKMVGRFFGREDSLLCATGFLATMSSICAVAKEGDLIVGDNRLHASLRSGM
CCHHHHHHHHHHHHHHCCCHHEEEECCCHHHHHHHHHHCCCCCEEEECHHHHHHHHHHH

KLSGAKEMLFRHNNWHHLQQTLAKHRRKYKNCWIVIESVYSMDGDIADLPVVRRLADQYN
HHCCCEEEEECCCCHHHHHHHHHHHHHHCCCEEEEEECEEECCCCCCCHHHHHHHHHHCC

CRILLDEAHGLGVLGKTGRGLEEHFNMPGAADVIVGTFSKSIGGVGGYITGDNDLVEFLD
CEEEEEHHCCCCEECCCCCCCHHHHCCCCCCCEEEEEEHHHHCCCCCCEEECCHHHHHHHH

FHAPGSVFSAPLTAYSAGGAMMAFELMQGEQSWRIAKAQENAKYLRHALQTGLGLWPKDY
HHCCCCEECCCCCHHCHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHCCCCCCCCC

PAERKFELEGVACTTVIPVVFPHDGDRVFRVTQAMLKRGWMVAAAAYPACPLNRPRIRVT
CCHCCCCCCCCCCCCCCEEEEECCCHHHHHHHHHHHHHCCCEEEEECCCCCCCCCCCEEEEE

ATAAYNQKMMDEFVKSLVEVTVECPPTDMLR
ECCCCCHHHHHHHHHHHHHHHHCCCCCCCC
```

Letter	Secondary Structure Feature
C	Random Coil
H	β Sheet
E	α Helix

E αB-Crystallin Sequences

E.1 Wild Type Sequence

MDIAIHHPWIRRPFFPFHSPSRLFDQFFGEHLLESDLFPTSTSLSPFYLRPPSFLRAPSW
FDTGLSEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVHGKHEERQDEHGFISREFHR
KYRIPADVDPLTITSSLSSDGVLTVNGPRKQVSGPERTIPITREEKPAVTAAPKK

E.2 Crystal Structure Sequence

GAMEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVHGKHEERQDEHGFISREFHRKYR
IPADVDPLTITSSLSSDGVLTVNGPRKQVSGPER

Truncation

MDIAIHHPWIRRPFFPFHSPSRLFDQFFGEHLLESDLFPTSTSLSPFYLRPPSFLRAPSW
FDTGLSEMRLEKDRFSVNLDVKHFSPEELKVKVLGDVIEVHGKHEERQDEHGFISREFHR
KYRIPADVDPLTITSSLSSDGVLTVNGPRKQVSGPERTIPITREEKPAVTAAPKK

Section	Colour
Residues not present in the crystal structure	Red

F Heat Shock Protein 16.5 Sequences

F.1 Wild Type Sequence

MFGRDPFDSLFERMFKEFFATPMTGTTMIQSSTGIQISGKGFMPISIIEGDQHIKVIAWL
PGVNKEDIILNAVGDTLEIRAKRSPLMITESERIIYSEIPEEEEIYRTIKLPATVKEENA
SAKFENGVLSVILPKAESSIKKGINIE

F.2 Truncation Mutations 3FR and 4FR

MFGRDPFDSLFERMFKEFFATPMTGTTMIQSSTGIQISGKGFMPISIIEGDQHIKVIAWL
PGVNKEDIILNAVGDTLEIRAKRSPLMITESERIIYSEIPEEEEIYRTIKLPATVKEENA
SAKFENGVLSVILPKAESSIKKGINIE

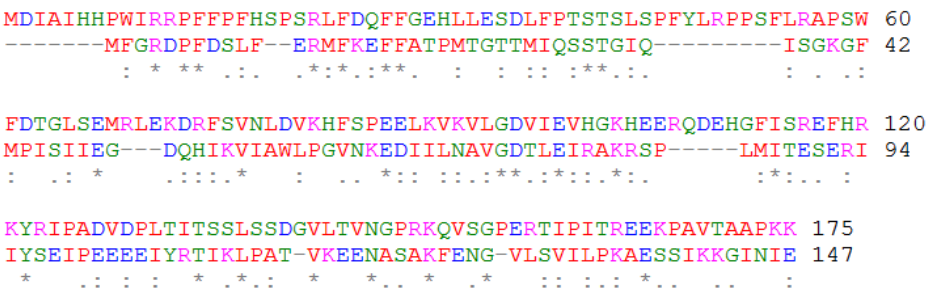
Section	Colour
Extent of 3FR Deletion	Red
Extent of 4FR Deletion	Blue

F.3 Point Mutation R107G

MFGRDPFDSLFERMFKEFFATPMTGTTMIQSSTGIQISGKGFMPISIIEGDQHIKVIAWL
PGVNKEDIILNAVGDTLEIRAKRSPLMITESERIIYSEIPEEEEIYGTIKLPATVKEENA
SAKFENGVLSVILPKAESSIKKGINIE

Section	Colour
Site of Arginine to Glycine Mutation	Green

G αB-Crystallin Sequence Alignment



Abbreviation	**Enzyme**
HsaBC	*Homo sapiens* αB-Crystallin
MjHSP	*Methanococcus jannaschii* Small Heat Shock Protein 16.5

Amino Acids	**Colour**	**Characteristics**
AVFPMILW	Red	Small (Small & Hydrophobic (Incl. Aromatic -Y))
DE	Blue	Acidic
RK	Magenta	Basic
STYHCNGQ	Green	Hydroxyl + Amine + Basic - Q

H *In vivo* Functionality Study Protocol Using Mass Spectrometry

1. Transform chemically competent *E. coli* (Rosetta, pLysS), plating out, and incubating overnight at 37℃;

2. Pick one colony, and inoculate a 2ml overnight culture (LB broth in a 15ml falcon tube, with the correct antibiotics), incubate at 37℃ with shaking at 120 rpm;

3. Inoculate a 15ml culture (Power Broth, in a 50ml falcon tube, with the correct antibiotics) with 150μl of the overnight culture;

4. Grow with shaking at 120 rpm at 37℃ until the OD 600nm reaches 0.6 and induce with IPTG to a final concentration of 1mM;

5. Incubate for 3 hours, then harvest the cells by centrifugation at 6000×g for 10 minutes;

6. Decant the supernatant and resuspend the pellet in 1ml 10:10:3 chloroform: methanol:water (HPLC grade reagents), reserving a small amount of the pellet for analysis by SDS-PAGE to check expression levels [N.B. From this point forward, all contact with plastic ware should be avoided where possible];

7. Vortex the resuspended materials with a small quantity of glass beads for 5 minutes to ensure cell lysis;

8. Spin the cell material at 14,000 for 10 minutes;

9. Using a fine-tipped Pasteur pipette carefully extract the lower organic phase, containing the desired lipids.

Colloquia & Conferences

Colloquia

14th October 2008 Prof Tom McLeish, University of Durham

"*Listening to Noise in Biological Physics from Protein Dynamics to Evolutionary Landscapes*"

15th October 2008 Prof Rob Field, John Innes Centre

"*From biofuels to homeland security: challenges for chemistry and chemical biology*"

22nd October 2008 Prof Mike Ward, University of Sheffield

"*Polyhedral coordination cages*"

11th November 2008 Prof Rick Lewis, University of Newcastle upon Tyne

"*Molecular architecture of the stressosome, a signal integration and transduction hub*"

12th November 2008 Prof Alan Bond, Monash University, Australia

"*Electrochemistry and photochemistry in ionic liquids: separating fact from fiction*"

25th November 2008 Dr Peter Urwin, University of Leeds

"*Pests, Trials and Tribulations*"

18th January 2009 Prof Mark Weller, Southampton University

"*From Greenland to Grenoble; The Structural Chemistry of New Frameworks and All Sorts of Hydrogen Compounds*"

21st January 2009 Dr Fred Manby, University of Bristol

"*Electronic structure theory for solids, liquids and enzymes*"

28th October 2009 Prof Bill Clegg, University of Newcastle

"*Bringing the power of syncrotron radiation to the chemical community*"

5th March 2009 Prof Jason Micklefield, University of Manchester

"*Synthetic Biology: From Nucleic acid Redesign to Engineering Biosynthetic Pathways*"

6th May 2009 Prof Erick M. Carreira, ETH Zurich, Switzerland

"*Discovery and Surprises with Small Molecules*"

7th May 2009 Prof Wilson Poon, University of Edinburgh

"*It's a Bug's Life: A Survey of The Physics of Bacteria*"

12th May 2009 Professor Gerard Meijer, Fritz-Haber Institute of the MPG, Berlin

"*Cold Molecules*"

12th May 2009 Dr Sean Munro, Cambridge University

"*How to find the Golgi apparatus - GTPases, coiled-coil proteins and transmembrane domains*"

Conferences

16th December 2008 Newcastle University

British Crystallographic Association Biological Structures Group Winter Meeting

3rd – 5th May 2009 University of Nottingham

CCP4 Study Weekend: "*Experimental Phasing and Radiation Damage*"

28th March – 6th April 2009 Trevelyan College, University of Durham

British Crystallographic Association Chemical Crystallography Group Intensive Teaching School in X-ray Structure Analysis

21st – 23rd April 2009 University of Loughborough

British Crystallographic Association Spring Meeting: "*Dynamic Crystallography*"

Presented poster: "Towards the Crystal Structure of the Serine Palmitoyltransferase from *T. gondii*"

6th – 8th January 2010 University of Nottingham

CCP4 Study Weekend: "*From Crystal to Structure with CCP4*"